UNIVERSITÉ DE TOURS

ÉCOLE DOCTORALE : SSBCV

E.A. 2106 « Biomolécules et Biotechnologies Végétales »

THÈSE présentée par :

Pamela Lemos Cruz

Soutenue le : **17 décembre 2021**

pour obtenir le grade de : **Docteur de l'université de Tours**

Discipline/ Spécialité : Santé - Sciences Biologiques et Chimie du Vivant

Élucidation et ingénierie de la biosynthèse d'alcaloïdes d'intérêt pharmaceutique originaires des Apocynacées

THÈSE dirigée par :

M. CLASTRE Marc	Professeur des Universités, Université de Tours

co-encadrée par :

M. COURDAVAULT Vincent	Maître de Conférences, Université de Tours

RAPPORTEURS :

Mme REMAUD-SIMEON Magali	Professeure des Universités, INSA de Toulouse
M. NAVROT Nicolas	Maître de Conférences (HDR), Université de Strasbourg

---

JURY :

Mme AUGE-GOUILLOU Corinne	Professeure des Universités, Université de Tours, présidente du jury
M. CLASTRE Marc	Professeur des Universités, Université de Tours
M. COURDAVAULT Vincent	Maître de Conférences, Université de Tours
M. HAWKINS Simon	Professeur des Universités, Université de Lille
M. NAVROT Nicolas	Maître de conférences (HDR), Université de Strasbourg
Mme REMAUD-SIMEON Magali	Professeure des Universités, INSA de Toulouse

Acknowledgements

I acknowledge Dr. Nathalie Guivarc'h, professor and director at EA2106 « Biomolécules et Biotechnologies Végétales » (BBV) at Tours for believing in me and for have made a huge effort in order to offer me the opportunity for coming to France and work in her laboratory. I am indebted to you for your support and for taking care of the research projects of the team.

I would like to express my deepest gratitude to my thesis director, Dr. Marc Clastre, professor at EA2106 « Biomolécules et Biotechnologies Végétales », whose sincerity and encouragement I will never forget. Marc has been an inspiration as I hurtled through the path of this PhD. I am thankful for the extraordinary experiences he arranged for me and for providing opportunities for me to grow professionally.

This work would not have been possible without the constant support and assistance of my co-supervisor Dr. Vincent Courdavault, assistant professor at EA2106 BBV. His patience, knowledge, and resilience is something that I will always keep aspiring to. Thanks Vincent for your continued guidance and an endless supply of fascinating projects. Even when I was feeling discouraged, you believed in my potential. Thank you for all the effort that you and Nathalie have done to help me accomplish this dream.

I would like to express my sincere gratitude to the professor Dr. Magali Remaud-Simeon and to the assistant professor Dr. Nicolas Navrot who accepted to be reviewers of this work, as well as the professor Dr. Simon Hawkins and the professor Dr. Corinne Auge-Gouillou, who gave me the honor of being part of the thesis committee.

I am grateful for having had a lot of help from other EA2106 EA2106 « Biomolécules et Biotechnologies Végétales » members in my research. Thank you very much Dr. Arnaud Lanoue, Dr. Sébastien Besseau, Dr. Audrey Oudin and Dr. Johan-Owen De Craene for all your advices and helping hands.

I would like to thank the post-doc Ines Carqueijeiro for all her support during all these years of work. Thank you Ines for training me and for helping me to get through this thesis. Very special thanks to all my other teammates who helped me in so many ways. Thank you Kostas, Natalja, Emily, Louis-Valentin, Gregory, Jennifer, Manon, Clément, Chaima, Thibaut, Kevin, Valentin, Damien and Marianne.

Most importantly, I am grateful to God and for my family and boyfriend's unconditional, unequivocal, and loving support. I love you all deeply.

Résumé

La pervenche de Madagascar (*Catharanthus roseus*) est une plante de la famille des Apocynacées produisant des alcaloïdes indoliques monoterpéniques (AIM) dont certains présentent un intérêt thérapeutique majeur. *C. roseus* est considérée aujourd'hui comme une plante médicinale modèle en devenant la référence en matière d'étude d'un métabolisme spécialisé, celui des AIM. L'avancée des connaissances sur ce métabolisme a été impulsée par les développements de ressources transcriptomiques alliés à des méthodes de caractérisation fonctionnelle des gènes. L'expertise acquise sur *C. roseus* a également permis la transposition de ces connaissances à d'autres plantes de la famille des Apocynacées. L'objectif de la thèse a consisté à étudier des enzymes de la famille des *O*-méthyltransférases du métabolisme alcaloïdique dans le but de les valoriser dans des approches d'ingénierie métabolique. Ainsi, suite à l'exploitation de données omiques, un gène codant une nouvelle isoforme enzymatique de la tabersonine-16-O-méthyltransférase (16OMT2) a été isolé chez *C. roseus* ainsi qu'un gène orthologue *Vm16OMT* chez l'espèce *Vinca minor*. La validation fonctionnelle du gène *16OMT2* a été réalisée par des études de pertes de fonction grâce à la technique VIGS (Virus-Induced Gene Silencing) et par des études d'expression transitoire utilisant la méthode d'agroinfiltration. Par ailleurs, les deux enzymes ont été caractérisées sur le plan biochimique et leur localisation subcellulaire déterminée. Des recherches en ingénierie métabolique de la levure *Saccharomyces cerevisiae* utilisant les gènes *16OMT2* et *Vm16OMT* ont été menées en vue d'optimiser les premières étapes de bioconversion de la tabersonine en vindoline (l'un des deux précurseurs des alcaloïdes à activité anticancéreuse). Au final, en exploitant les résultats obtenus avec le gène *16OMT2* et d'autres résultats antérieurs obtenus par l'équipe, ces travaux ont conduit à produire en bioréacteur 266 mg/L de vindoline ouvrant la voie à une future production industrielle de ce métabolite.

Mots-clefs: Apocynaceae, *Catharanthus roseus*, alcaloïdes indoliques monoterpéniques, biosynthèse, ingénierie métabolique.

Summary

The Madagascar periwinkle (*Catharanthus roseus*) is a model medicinal plant of the Apocynaceae family producing indole monoterpene alkaloids (MIA) which became the reference in the study of specialized metabolism. The increased knowledge on this metabolism has been driven by the development of transcriptomic resources combined with new methods of functional characterization of genes, which also allowed the transposition of this knowledge to other plants of the Apocynaceae family. The objective of the thesis was to study the *O*-methyltransferase family enzymes involved in the alkaloid metabolism, in order to develop them in metabolic engineering approaches. Thus, following the study of the omics data, a gene encoding a new enzyme isoform of tabersonine-16-*O*-methyltransferase (16OMT2) was isolated in *C. roseus*, as well as an ortholog gene *Vm16OMT* in the *Vinca minor* species. Functional validation of the *16OMT2* gene was carried out by loss-of-function studies using the VIGS (Virus-Induced Gene Silencing) technique and by transient expression studies using the agroinfiltration method. In addition, the two enzymes were characterized biochemically and their subcellular localization described. The metabolic engineering of the yeast *Saccharomyces cerevisiae* was carried out using the *16OMT2* and *Vm16OMT* genes in order to optimize the first steps of tabersonine to vindoline bioconversion (one of the two precursors of alkaloids with anticancer activity). Finally, by exploiting the results obtained with the *16OMT2* gene and other previous results obtained by the team, this work led to the production of 266 mg/L of vindoline in a bioreactor, paving the way for future industrial production of this metabolite.

Mots-clefs: Apocynaceae, *Catharanthus roseus*, monoterpene indole alkaloids, biosynthesis, metabolic engineering.

Abbreviations

16OMT : hydroxytabersonine 16-O-méthyltransferase
16OMT2: hydroxytabersonine 16-O-méthyltransférase 2
AA : amino acid
Acetyl-CoA: acetyl-coenzyme A
BIA: benzylisoquinoline alkaloids
BiFC : bimolecular fluorescence complementation
CCoAOMT: caffeoyl coenzyme A *O*-methyltransferase
cDNA: complementary deoxyribonucleic acid
CO_2: carbon dioxide
COMT: caffeic acid *O*-methyltransferase
CPR : NADPH cytochrome P450 reductase
CS : catharanthine synthase
CYP, **P450** or **CYP450**: cytochrome P450
CYP450-reductase: cytochrome P450 reductase
CYPD6: cytochrome P450 2D6
D4H : desacetoxyvindoline 4-hydroxylase
DAT: deacetylvindoline 4-O-acetyltransferase
DMAPP : dimethylallyl diphosphate
DNA: deoxyribonucleic acid
ER: endoplasmic reticulum
G10H : geraniol 10-hydroxylase
GFP : green fluorescent protein
GO: geissoschizine oxydase
GPP : geranyl diphosphate
GPPS : geranyl diphosphate synthase
GS: geissoschizine synthase
HMBPP: 4-hydroxy-3-methylbut-2-enyl diphosphate
HYS: heteroyohimbine synthase
IPAP: internal phloem associated parenchyma
IPP : isopentenyl diphosphate
MECPP: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate
MEP: 2-C-methyl-D-erythritol-4-phosphate
MIA: monoterpene indole alkaloid
mRNA: messenger ribonucleic acid
MVA: mevalonic acid
MVD : mevalonate 5-diphosphate decarboxylase
MVK : mevalonate kinase
NADH : nicotinamide adenine dinucleotide
NADPH : nicotinamide adenine dinucleotide phosphate
NMT: 3-hydroxy-16-methoxy-2,3-dihydrotabersonine N-methyltransferase
OE: overexpression
OMT: 16-*O*-methyltransferase
PAS: precondylocarpine acetate synthase
PPM: party per million
PRX1 : peroxydase 1
REDOX1: cinnamyl alcohol dehydrogenase-like enzyme

REDOX2: aldoketo–type reductase
RNA: ribonucleic acid
RNA-seq: RNA sequencing
SAH: S-adenosyl-L-homocysteine
SAM: S-adenosyl-L-methionine
SAM2: S-adenosyl-L-methionine synthase 2
SAT: stemmadenine-O-acetyltransferase
SDGs: Sustainable Development Goals
SDR: short-chain dehydrogenase/reductase
SGD : strictosidine β-glucosidase
shSGD: short strictosidine β-glucosidase
SMs: specialized metabolites
STR : strictosidine synthase
T16H : tabersonine 16-hydroxylase
T16H1 : tabersonine 16-hydroxylase 1
T16H2 : tabersonine 16-hydroxylase 2
T19H: tabersonine 19-hydroxylase
T3O: tabersonine-3-oxidase
T3R: tabersonine-3-reductase
TAT: tabersonine 19-*O*-acetyltransferase
TDC: tryptophan decarboxylase
THAS : tetrahydroalstonine synthase
TRL: 2,3-dihydro-3-hydroxytabersonine-N-methyl transferase
TRV: rattle tobacco virus
TS: tabersonine synthase
UDP-glucose: uridine diphosphate glucose
UV : ultra-violet
VIGS: virus induced gene silencing
VmOMT16: vincadifformine-16-*O*-methyltransferase

List of Contents

I. Introduction 1

1. General introduction 2

1.1. Plant specialized metabolites 2

1.2. Flavonoids 4

1.3. Alkaloids 5

2. The evolution of *O*-Methyltransferases: a review 7

2.1. Introduction 7

2.2. Methyltransferases 10

2.3. The O-methyltransferases superfamily 13

2.3.1. Phenylpropanoids OMTs 14

2.3.2. Flavonoid OMTs 15

2.3.3. Alkaloid OMTs 16

2.4. Evolution of new functionalities in OMTs 19

2.5. Conclusion 22

3. Medicinal alkaloids 23

4. *Catharanthus roseus* and the monoterpene indole alkaloid biosynthesis 26

4.1. *C.roseus* MIAs 26

4.2. *C.roseus* MIA biosynthesis 28

4.2.1. Enzyme promiscuity: T3O, T3R, NMT, DAT and D4H 32

4.2.2. Protein isoforms 34

4.3. Role of MIAs in *C. roseus* metabolism 35

5. Metabolic engineering 38

5.1. Metabolic engineering applications 38

5.2. The elucidation of plants natural product biosynthetic pathways 39

5.3. Improving the mechanisms of gene transfer into the heterologous hosts 41

5.4. Optimization of the biosynthetic pathways up to an industrial level 42

5.5. Expand the library of molecules by the creation of non-natural functional products 43

II.Thesis organisation and objectives 45

III.Results 50

Part 1 : Methods used for the identification and validation of biosynthetic genes 51

Article 1: A biolistic-mediated virus-induced gene silencing in Apocynaceae to map biosynthetic pathways of alkaloids 53

Article 2: A rapid and efficient vacuum-based agroinfiltration protocol for transient gene overexpression in leaves of *Catharanthus roseus* 73

Part 2: 16-O-methyltransferase identification and involvement in alkaloids biosynthesis 95

Article 3 : Vindoline synthesis is coordinated by two 16-hydroxytabersonine-O-methyltransferase isoforms in leaves of *Catharanthus roseus* 96

Article 4 : Identifying genes involved in alkaloid biosynthesis in *Vinca minor* through transcriptomics and gene co-expression analysis 153

Part 3: Metabolic engineering of the vindoline biosynthetic pathway in yeast..... 223

Article 5 : Optimization of tabersonine methoxylation to increase vindoline precursor synthesis in yeast cell factories..... 224

Article 6 : Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories..... 246

IV. Conclusion..... 271

1. Gene validation: methods and considerations..... 274
2. Characterisation of isoformes and orthologues of *C. roseus* MIA genes 275
3. Production of vindoline by yeast metabolic engineering..... 278

V. Perspectives..... 282

VI. References 289

List of figures

Figure 1: Representative plant specialized metabolites of high socio-economic impact. 3
Figure 2: Reactions involved in the biosynthesis of vinblastine. 6
Figure 3: Natural products often found in plants.....8
Figure 4: Most common mechanisms of chemical decorations in plants..... 9
Figure 5: Reaction catalyzed by the tabersonine 16-*O*-methyltransferase (16OMT) on the Vindoline's pathway of *C. roseus.*.....10
Figure 6: Biosynthesis of the three building blocks of lignin in a simplified model. 15
Figure 7: Chemical structure of the natural product morphine.. 19
Figure 8: Phylogenetic relationships among representative functionally characterized plant O-methyltransferases.....20
Figure 9: *C. roseus* monoterpene indole alkaloids. 26
Figure 10: A summarized view of MIA biosynthesis in *C. roseus.* 28
Figure 11: *C. roseus* MIA biosynthetic pathway: An integrated model of its cellular and subcellular architecture. 31
Figure 12: Tabersonine fate in *C. roseus* leaves and its tissue distribution. 33
Figure 13: Deglycosylation of strictosidine and associated reactions in *C. roseus*.....36
Figure 14: A standard procedure for heterologous organism production of natural products. 39
Figure 15: The use of high-throughput transcriptomic data for discovery of MIA biosynthetic genes. 41
Figure 16: Expansionof the library of molecules by associating different *C. roseus* enzymes. ..44
Figure 17: From gene identification to production of natural products in yeast. 273
Figure 18: Updated Phylogenetic tree with the genes studied in this thesis. 277
Figure 19: Optimization of the pathway for vindoline production.....280

List of tables

Table 1: Typical methyltransferase in plants, their methyl acceptor molecules, and their functional significance.....12
Table 2: Representative alkaloid plant *O*-methyltransferases and their biological significance..17
Table 3: Representative natural medicinal alkaloids.....24

I. Introduction

1. General introduction

1.1. Plant specialized metabolites

Plants are responsible for the assembly of a vast collection of molecules that are incredibly diverse chemically and arise from both primary and secondary metabolisms. Primary metabolism is associated with vital functions of the plant, such as growth and development, and are present in every living cell (Bhambhani et al., 2021). Conversely, secondary metabolites, or specialized metabolites (SMs), are cell-specific and their presence and localization are correlated to a required specific function (Pott et al., 2019). The expansion of the knowledge associated with SMs led to the discovery that these molecules date from 500 million years ago which corresponds to the beginning of the land colonization by plants. This new environment brought new abiotic and biotic stress to the plants, including new predators. Surprisingly, these unique metabolites sustained the adaptation of plants by providing ultraviolet protection, interspecies interactions and anti-microorganism barriers (Banks et al., 2011; Kusano et al., 2019). In addition, defense response signaling, innate immunity and direct response to environmental both biotic and abiotic stress conditions were also related to the SMs (Wink, 2008; Yang et al., 2018).

Chemically, the SMs are a huge family in which more than 20 thousand molecules are found. They can be roughly divided into three superfamilies of chemical compounds: phenolics, terpenoids and alkaloids (Bourgaud et al., 2001). The phenolic compounds are characterized by the presence of a benzene ring carrying at least one hydroxyl group and the most abundant members of this superfamily are the phenylpropenes, the coumarins, the lignans, flavonoids and the tannins. On the other hand, the terpenes are derived from the assembly of isoprene units containing 5 carbon atoms (C5), being also found in bigger chains as mono- (C10), sesqui- (C15), di- (C20), tri- (C30), tetra- (C40), and polyterpenes (>40). Finally, the alkaloids are structures containing a nitrogen atom or atoms (amino or amido in some cases), usually inside a ring (cyclic) system (Kurek, 2019).

Nature has always been the source of human medicines and it was with the use of plant extracts that humans could find the cure for their diseases. The pharmacological activity of many plants are attributed to simple phenolics possessing antimicrobial and diuretic effects (*Arctostaphylos uva-ursi, Vaccinium macrocarpon)* (Yarnell, 2002); or to the flavonoids and their anti-inflammatory and antiallergic effects (*Calendula officinalis, Passiflora edulis)* (Montanher

et al., 2007)*;* and specially to the antineoplastic, hypertensive and hypotensive effects of the alkaloids (*Catharanthus roseus, Papaver somniferum)* (Noble, 1990; Stermitz, 2000). Consequently, the SMs have become an important part of active traditional medicine and, today, are present in innumerous health, cosmetic, recreational and food products, which considerably increased the commercial importance and value of plants (figure 1). On the next topic, we'll discuss the two classes of SMs that are closely related and that will help the understanding of some discussion throughout this thesis, the plant flavonoids and alkaloids.

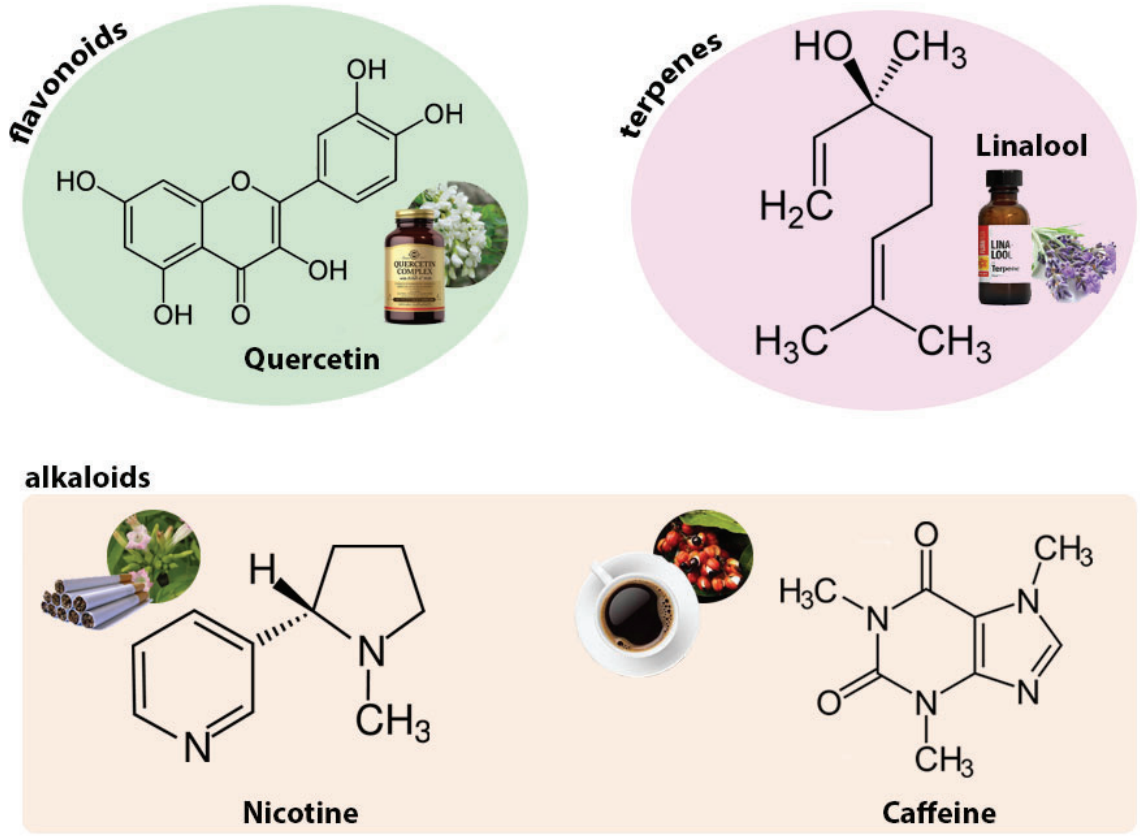


Figure 1: Representative plant specialized metabolites of high socio-economic impact. Flavonoids: Quercetin is widely used due to its antioxidant, anti-inflammatory and antihistaminic properties and its sourced from the onions, grapes, berries, cherries, broccoli, and citrus fruits. Terpenes: Linalool is sourced from lamiaceae (mints), lauraceae (laurels, cinnamon, rosewood) and rutaceae (citrus) and it is known for its relaxation properties that fight anxiety and insomnia. Alkaloids: Nicotine is sourced from tobacco species and is mostly used on the production of cigarettes. Caffeine is produced from cocoa beans, kola nuts, tea leaves and coffee beans and it is used to produce two of the most common drinks, the coffee and the tea.

1.2. Flavonoids

Flavonoids are the largest class of polyphenols, including around 9000 molecules. The most common molecules in this class are the chalcones, flavones, flavanols, anthocyanins and proanthocyanidins (Yonekura-Sakakibara & Hanada, 2011). These molecules are vastly found in the plant kingdom, including angiosperms, gymnosperms, and pteridophytes (Wen et al., 2020). In the model plant *Arabidopsis thaliana*, for example, around 54 different flavonoids molecules are found, between flavanols, anthocyanins and proanthocyanidins (Saito et al., 2013). This large diversity inside the flavonoids group is the source of their huge variety of functions *in planta.* The most important role of flavonoids is to display ultraviolet-B (UV-B, 280-315 nm) protection to leaves, which correlates these molecules to the plants ability of survival in deteriorating environments (Peng et al., 2017; Tohge et al., 2016). In addition, this protective function is also effective against other abiotic stress such as salt and drought and against biotic stress such as pathogen and herbivore attack, being crucial to plant defense (Khalid et al., 2019). Other flavonoids roles include lateral root development (Brown et al., 2001), maximal gravitropic response (Buer & Muday, 2004), and regulation of this response to ethylene (Buer et al., 2006).

Chemically, flavonoids are derived from the condensation of the amino acid phenylalanine with malonyl-CoA, which results in a common diphenyl propane backbone (C6-C3-C6) with two aromatic rings, linked by a three-carbon chain (Saito et al., 2013; Tohge & Fernie, 2017). The final structure will consist of a benzene ring linked to a pyrone ring and finally, a phenyl ring located at the position 2 or 3. The sub-classification of each flavonoid will depend on the presence and the position of their side structures, being classified as flavones, flavonols, flavanones, flavanols, anthocyanins or isoflavones. For example, flavones present a hydroxyl group (-OH) on the position 5 and 7 of its benzene ring, while flavanones possess the same hydroxyl group only on the position 5.

Flavonoids also display several properties that are indispensable for human health, such as antioxidant, antiviral, antibacterial, antifungal, anti-inflammatory and anticancer (Bojić et al., 2019; Grigalius & Petrikaite, 2017; Khalid et al., 2019). Appropriately, these molecules became an indispensable component of nutraceutical, cosmetic and medicinal products. Notably, chemical decorations of flavonoids are associated with important changes in its solubility, storage and biological activities, making these molecules targets of natural or non-natural chemical decorations such as methylation, which is proved to considerably increase their biological activity

(Saito et al., 2013). More information about how methylation can alter a flavonoid biological activity will be presented in the chapter II.

1.3. Alkaloids

The alkaloids are a large group of molecules and represent around 20% of plant-based SMs across the plant kingdom (Dey et al., 2020). The main characteristic of these molecules is the typical presence of a nitrogen atom in a heterocyclic ring. Alkaloids are extremely diverse, whether chemically, biochemically or physiologically. Different arrangements of alkaloids are found in nature, leading to different functional entities which are responsible for a wide range of biological properties for these molecules (Dey et al., 2020; Kurek, 2019). Based on their heterocyclic ring system and biosynthetic precursor, alkaloids are classified into diverse categories including indole, purine, quinoline, isoquinoline, tropane, imidazole, among others (Hussein & A. El-Anssary, 2019; Dey et al., 2020). Under the classification upon their biosynthetic origin, four large groups of alkaloids can be described in plants:

1) Monoterpene indole alkaloids: Derived from the condensation of the amino acid tryptophan with the monoterpenic derivative secologanin
2) Benzyliquinoline alkaloids: Derived from tyrosine
3) Tropane alkaloids: Derived from arginine and phenylalanine
4) Purine alkaloids: Derived from purine

Further classifications of alkaloids will be related to the different possible chemical decorations that the molecule will overcome. On the figure 2, a complex dimeric monoterpene alkaloid from *C. roseus* is represented. Its base skeleton is issue from the condensation of tryptophan and secologanine, being classified as a mnoterpene indole alkaloid, followed by multiple chemical decorations such as methylation, oxidation, reduction, glycosylation, acetylation, hydroxylation, and epoxidation, which are responsible for its biological functionalities.

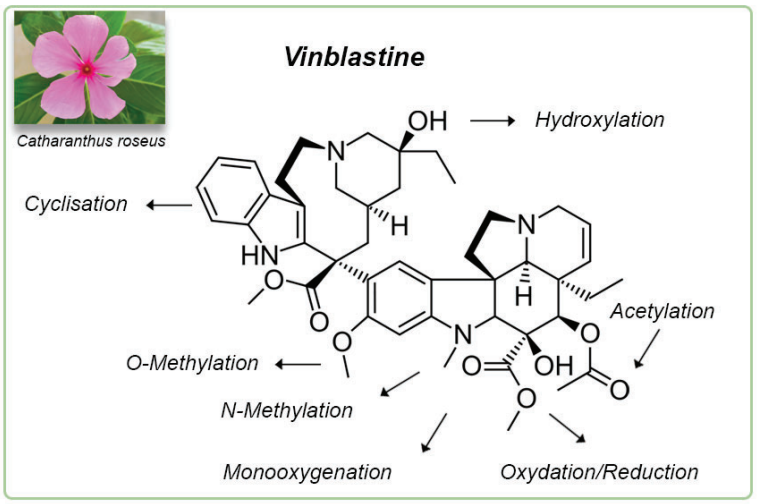


Figure 2: Reactions involved in the biosynthesis of vinblastine.

The role of alkaloids *in planta* is related to the storage of nitrogen and defense against predators, especially animals and insects due to their general toxicity and deterrence capacity (Matsuura & Fett-Neto, 2015). In addition, studies have shown that some alkaloids also play a role in plant growth due to their structure similarity with main known plant regulators. This allelopathic effect was observed in cocaine from *Erythroxylon coca* Lamk and also with berberine, sanguinarine, and gramine, where these alkaloids inhibited the seedling growth of *Lactuca sativa* and *Lepidium sativum* (Jogaiah & Abdelrahman, 2019; Matsuura & Fett-Neto, 2015; Roberts, 1998). As flavonoids, alkaloids can also become more soluble or have an increased biological activity after chemical decorations such as hydroxylation and methylation, being a target for synthetic biology in order to build new active molecules (Carqueijeiro et al., 2018). Due to the importance of methylation of specialized metabolites, a review about O-methyltransferases is presented in the next chapter.

2. The evolution of O-Methyltransferases: a review

This chapter is part of a review article that I wrote during my thesis and since it was not yet finished and submitted to a scientific journal, an adapted version of the article manuscript is presented.

2.1. Introduction

Apart from animals, land plants developed singular features to survive in the environment. Without mobility, the synthetization of different organic compounds became their specialty. To date, more than 200,000 natural products have been described with a huge variety of chemical structures. This structural diversity of compounds are originated from different biosynthetic pathways, that can be described as multi-step, enzyme catalyzed process where simple substrates are converted into more complex products that display different functions in the plant, ranging from structural roles, defense against predators, until color and scent (O'Connor, 2012; Pickens et al., 2011; Yun et al., 2012).

Natural products are divided in two main groups, the primary and the specialized metabolites. Primary metabolites are organic molecules, which are essential for the growth and development of the plant, being mainly represented by core building molecules as nucleic acids, amino acids, sugars, and fatty acids; without these compounds, essential macromolecules as DNA, proteins and carbohydrates cannot be produced. On the other hand, specialized metabolites are not essential for the plant, because they are mainly involved with competitiveness and interaction with other organisms (Dugé De Bernonville et al., 2017; Takshak & Agrawal, 2019). These molecules are represented by a vast range of structures: from simple phenylpropanoids, passing through flavonoids and arriving at the most complex alkaloids (Li et al., 2020). Undoubtedly, the evolution of the secondary metabolites provided essential elements for the plants adaptation on the terrestrial environment and not solely for plants. These molecules became vital for humans as well, considering its extensive use by the pharmaceutical, food and cosmetics industry (Favarin et al., 2013; Grynkiewicz & Gadzikowska, 2008).

Starting from the simple structures of the primary metabolism, the structure of specialized metabolites can be divided into three big groups (figure 3): Phenolics, which are characterized by its aromatic ring with at least one hydroxyl group (—OH) bonded directly to it; Terpenes, which are prevenient from the assembly of five units of carbon isoprene units (C5); and finally, the

nitrogen-based compounds, as alkaloids for example. At a first look at these compounds, it is possible to recognize the chemical reactions possibly involved in its biosynthesis such as hydroxylation, cyclisation, oxidation, N and O-methylation and acylation. In addition, the decarboxylation and prenylation are also reactions that occur very often in natural products.

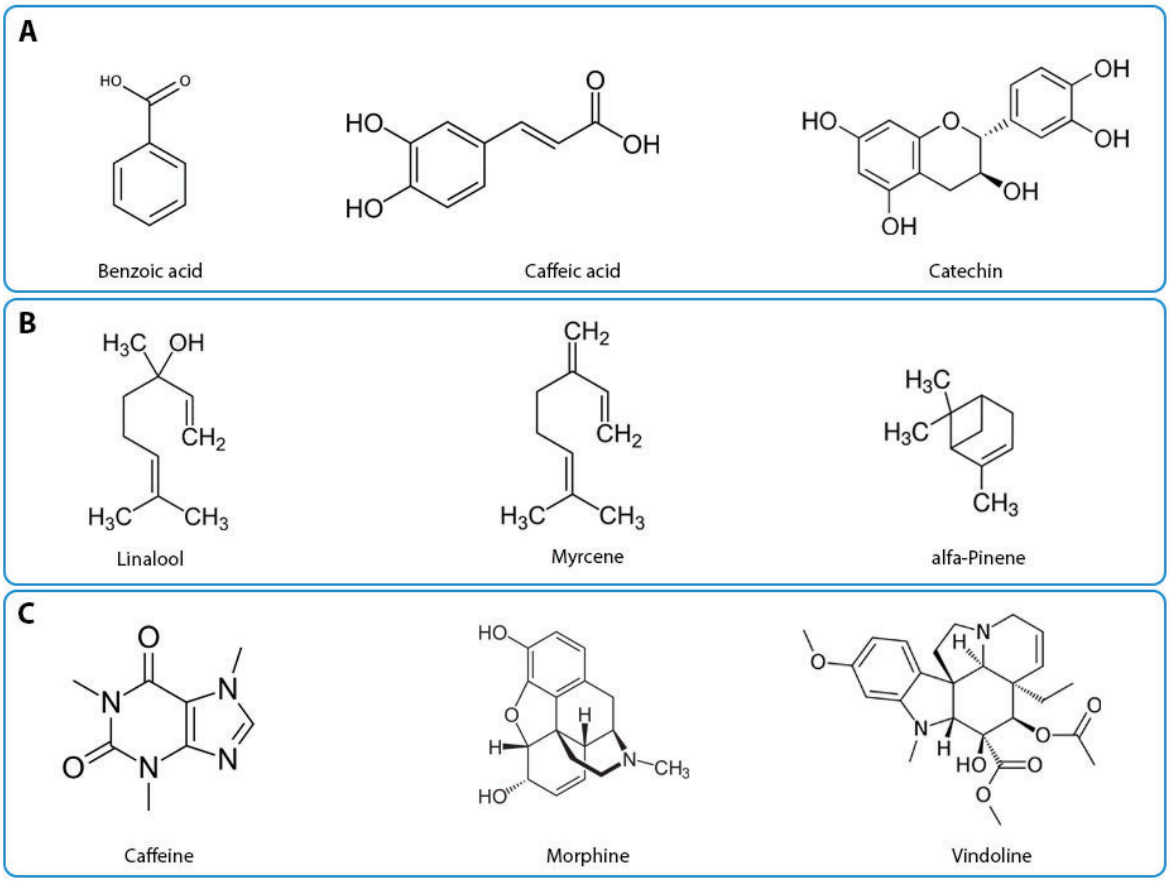


Figure 3: Natural products often found in plants. A: Phenolics. B: Terpenes and C: Alkaloids.

Due to the energetic cost of these chemical decorations, biological catalyzers are needed to promote these costly reactions. These catalyzers are called enzymes and are highly specialized, being classified according to their function. For example, the transfer of an oxygen molecule (O_2) to a substrate is catalyzed by mono (EC 1.14) or dioxygenases (EC 1.13), depending on the number of atoms of oxygen that are added into the substrates (figure 4). The source of oxygen used by these proteins is, on the majority, the dioxygen (O_2) and NADPH (nicotinamide adenine dinucleotide phosphate) as the hydrogen donor. The monooxygenases can be classified as non-heme or heme (P450s group) and minor classifications according to the substrate specificity are done due to major differences related to mechanisms of reception of the oxygen molecule and the use or not of redox partners (electron transfer between reductases and hydroxylases) (Holland & Weber, 2000).

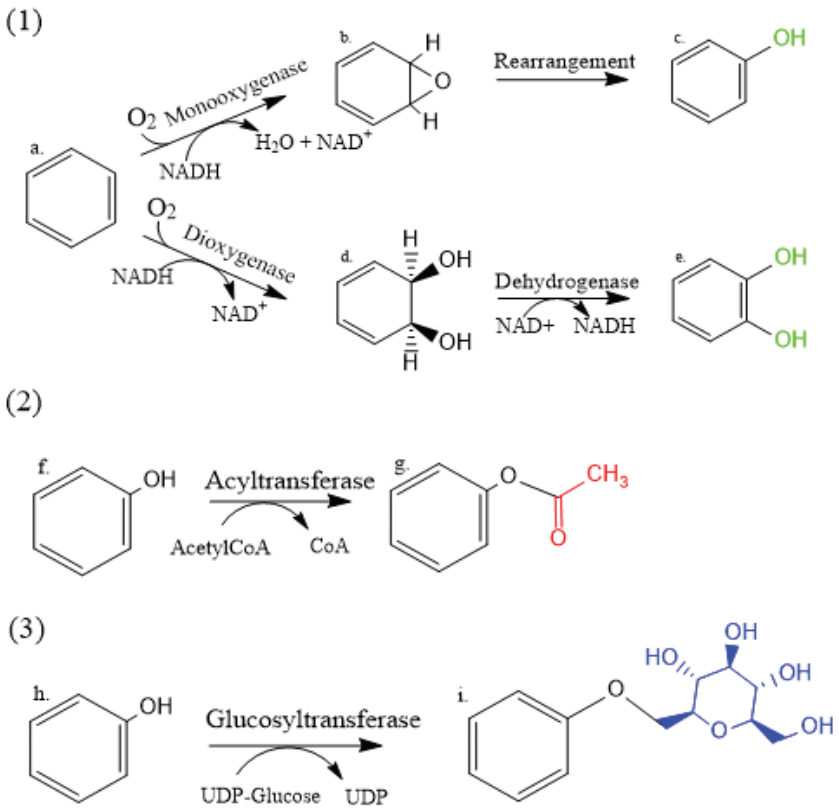


Figure 4: Most common mechanisms of chemical decorations in plants. (1) Hydroxylation by either mono or dioxygenases. (a.) aromatic substrate (b.) epoxide intermediate (c.) phenolic product (d.) *cis*-dihydriol (e.) catecholic product. (2) Acylation. (f.) phenolic substrate (g.) ketone (3) Glucosylation (h.) phenolic substrate (i) D-glucose insertion. This figure was inspired by (Ullrich & Hofrichter, 2007).

Another important class of proteins are the transferases (EC 2) which function is overall the transfer of groups from a donor to a receptor. Transferases are involved in hundreds of different biochemical pathways being essential to every living organism. An important example of a transferase subclass are the acyltransferases (EC 2.3), which catalyzes the reaction of acylation, meaning the transfer of an acyl group (R-CO) from a donor molecule to the receptor one (figure 4). Due to the high reactivity of the acyl group, this radical can interact with a range of molecules, for example a fatty acid or a hydroxycinnamic acid. Depending on the radical R of the acyl group, the reaction will be named differently. The most common acylation is the acetylation, which has a methyl radical -CH3 participation. This process occurs on account of the sole acetyl donor, the acetyl-coenzyme A (acetyl-CoA). Interestingly, this donor molecule participates in a big variety of functions including lipid synthesis and the carrying of glycolytic pyruvate to the tricarboxylic acid cycle. The structure of the acetyl-CoA consist of a membrane-

impermeant molecule with an acetyl moiety (-CH3CO) linked to the coenzyme A, which is an extremely rich in energy bond, allowing the transfer of the acetyl moiety to a vast array of acceptor, from simple molecules to proteins (Pietrocola et al., 2015).

Transferases act chemically decorating a molecule in order to regulate their transport or even to control their toxicity, which is one of the functions of the glucosyltransferases (EC 2.4) (Figure 4). The conversion of a lipophilic molecule into a hydrophilic one allows the solubilization of certain molecules, promoting their activation by making them bioavailable. The synthesis of these glucosides is achieved by the transfer of a glucose moiety from a nucleotide activated sugar donor molecule (UDP-glucose) to the aglycone (Iwakiri et al., 2013).

2.2. Methyltransferases

The most common class of transferases in plants are the methyltransferases (EC 2.1.1), which are responsible for the transfer of a methyl group -CH_3 from the donor S-adenosyl-L-methionine (SAM) to a receptor, possibly a carbon (C), a nitrogen (N), an oxygen (O), a sulfur (S), or even a halide, resulting on the formation of a methyl derivative and S-adenosyl-L-homocysteine (SAH) as products (figure 5). When the receptor of the methyl radical is a nitrogen, the enzyme will be called a *N*-methyltransferase, and when it is an oxygen, the enzyme which catalyze this reaction is called a *O*-methyltransferase (OMT) (Ibrahim & Muzac, 2000).

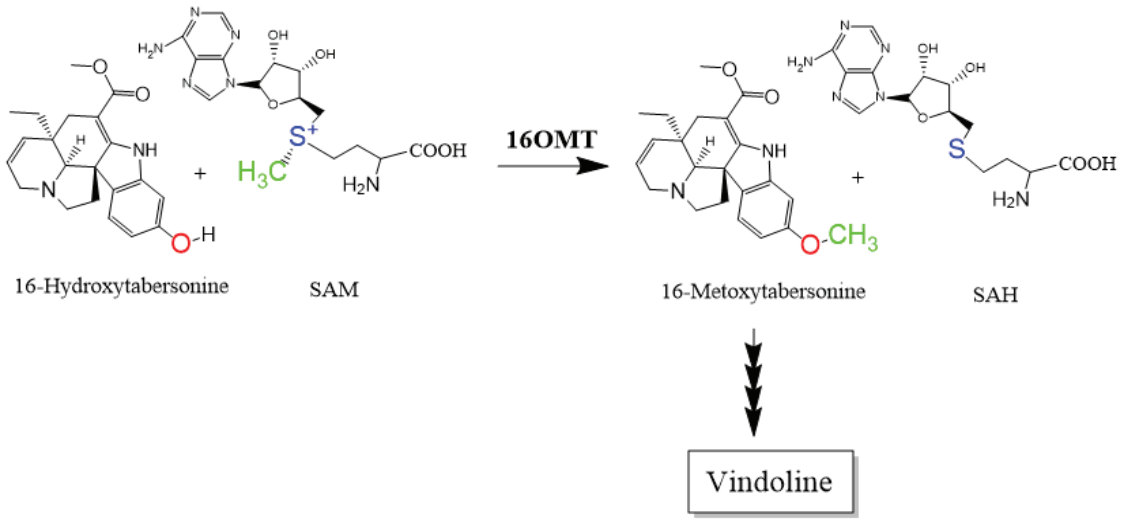


Figure 5: Reaction catalyzed by the tabersonine 16-*O*-methyltransferase (16OMT) on the Vindoline's pathway of *C. roseus*.

Despite of the simplicity of this reaction, most of these enzymes are very specific, not just in respect of stereochemistry of the substrate, but as well the exact position of the group to be methylated, resulting in very specialized biosynthetic pathways to supply variable needs of the plant. That is the reason why each simple reaction as methylation has such great importance on the production of natural products. The table 1 shows how the methyltransferases were first classified according to their substrate specificity.

As shown on the table 1, the OMTs group consists of the most important group of methyltransferases, being able to methylate several catecholic compounds as phenylpropanoids (subgroup A), flavonoids (subgroup B) and alkaloids (subgroup C). Overall, for the groups A, B and C, the methylation of an oxygen atom occurs within a hydroxyl group (-OH) leading to the reduction of its reactivity, changing the solubility of the molecule as well as, is some cases, its intracellular compartmentation (Ibrahim et al., 1998). From plants to animals, the OMTs are involved in several physiological reactions. In plants, OMTs have been described as responsible for an ample number of reactions, including the decoration of phenolic compounds that are involved in plant growth and development (Hahlbrock & Scheel, 1989); lignin biosynthesis (Lewis & Yamamoto, 1990); flavonoids and alkaloids that play essential role in plant defense (Dugé De Bernonville et al., 2017; Ibrahim et al., 1987).

Nowadays, with the progression of transcriptomics, the expansion of the knowledge of new OMTs brought to light several new features of these enzymes. Some OMTs are highly specific for a certain substrate (Levac et al., 2008; Nomura & Kutchan, 2010) and others can methylate a big variety of molecules (Cacace et al., 2003; Salim et al., 2018). This review will present important characteristics of OMTs, including their structure and function, but also will contribute to the understanding of the evolution process of this superfamily and the biotechnological perspective on this subject.

Table 1: Typical methyltransferases in plants, their methyl acceptor molecules, and their functional significance.

Methyltransferase/ Methyl acceptor molecule	Source	Functional role	Reference
O-Methyltransferases			
A) Phenylpropanoids			
Caffeic acid	*Several spp.*	Lignification	(Barakat et al., 2011; Ibdah et al., 2003; Lewis & Yamamoto, 1990)
Caffeoyl/CoA esters	*Several spp.*	Lignification	(Barakat et al., 2011; Ibdah et al., 2003; Lewis & Yamamoto, 1990)
B) Flavonoids			
Myricetin	*C. roseus*	Flavonol glycosides/ anthocyanins	(Schröder et al., 2004)
Daidzein	*M. sativa*	Alfafa phytoalexin	(Liu et al., 2006)
C) Alkaloids			
16-Hydroxytabersonine	*C. roseus*	Plant defense	(Levac et al., 2008)
Norcoclaurine	*Thalictrum tuberosum*	Berberine biosynthesis	(Morishige et al., 2010)
D) Polyketides			
Myo-inositol	*Mesembryanthemum crystallinum Porteresia coarctata Roxb. (Tateoka)*	Stress perception and amelioration	(Ishitani et al., 1996; Sengupta et al., 2008)
E) Peptides			
Peptide-L-isoaspartyl	*A. thaliana*	Damaged protein repair	(Byeon et al., 2016)
C-Methyltransferases			
Gamma-Tocopherol	*A. thaliana*	Tocopherol synthesis	(Liscombe et al., 2010)
N-Methyltransferases			
Putrescine	*Nicotiana tabacum*	Nicotine biosynthesis	(Khalil et al., 2015)
Rubisco-LSU	*N. tabacum, P. sativa*	Post-translational modification	(Hallard et al., 1997)
S-Methyltransferases			
Betaine/ homocysteine	*A. thaliana*	Methionine biosynthesis	(Byeon et al., 2014)
Halide methyltransferases			
Cl-, Br-, I-, [SH]-, [SCN]-	*R. sativus (daikon radish)*	Sulfur metabolism	(Itoh et al., 2009)

2.3. The O-methyltransferases superfamily

More recently, a new classification of plant OMTs has been done based on sequence similarity and signatures motifs. This classification resulted in two major groups PL-OMTI and PL-OMTII. The first group (PL-OMTI) consists of caffeoyl coenzyme A *O*-methyltransferases, (CCoAOMTs), which are in general small proteins (26-29 kDa) and mostly require Mg^{2+} as a cofactor. The second group (PL-OMTII) are represented by caffeic acid *O*-methyltransferase (COMTs), which are bigger proteins (38 kDa to 43 kDa) and can perform activity in the absence of any metal cations. On this second group are found the big majority of OMTs, including the ones that use (iso)flavonoids as substrate, benzylisoquinoline alkaloids (BIA), phenylpropanoids including hydroxycinnamic acids, and phenylpropenes (Lam et al., 2007).

An important characteristic of this superfamily is a conserved motif composed by five residues whose function was associated as folding sites of the SAM's pocket binding, the common methyl donor of a large variety of methyltransferases in plants. SAM dependent OMTs were reported on both PL-OMTI and PL-OMTII groups (Joshi & Chiang, 1998). Another interesting feature of OMTs is that most of these proteins form homodimers in solution. On the structural level, the binding between two OMT monomeric subunits occurs via around 150 from 350 residues total domain and is composed primarily of intertwined helices (Morris & Facchini, 2019). These helices interact with the same helices of the second monomer, covering around 22% of total protein surface area. Surprisingly, a study reported the presence of a trimer of CjSOMT (*Coptis japonica* (S)-scoulerine-9-*O*-methyltransferase) by gel filtration, however, it is unclear how this feature could relate to the protein sequence, since the authors found only a 20-amino acid extension of the N-terminus of the proteins, which normally is not involved in the dimerization domain (Morishige et al., 2000). Notably, homodimers are also not exclusive in dimers organization of OMTs, several studies suggest that heterodimerization (different monomers) may represent a key role in alkaloid biosynthesis (Li & Smolke, 2016; Park et al., 2018). Recently, an unknow mechanism of OMT heterodimers was described in *Papaver somniferum* in which a reported OMT heterodimer catalyzes a reaction not performed by either of each of its monomeric subunits (Park et al., 2018).

2.3.1. Phenylpropanoids OMTs

Plenty of biosynthetic pathways have been described to date within numerous species of plants in which *O*-methyltransferases play important roles. Lignin, for example, is a well-known natural compound due to its structural function in plants and also for the economics of food, paper and biorefining industry (Simon et al., 2018). Briefly, lignin is composed by three monolignols p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) which are arranged through several linkages in three-dimensional network (Patil et al., 2016) (figure 6). The pathway that generates these three molecules starts with an essential amino acid synthesized at the shikimate's pathway named phenylalanine, which suffers sequenced chemical decorations until the final product is obtained. The lignin pathway is an interesting starting point of study as it presents several close methylation reactions which are performed by methyltransferases from both I and II groups.

The first methylation occurring on the pathway is performed by a COMT (Caffeic Acid OMT that belong to the PL-OMTII group), which can perform both methylations of caffeic acid and 5-hydroxyferulic acid (figure 6, gray pathway). This protein is considered a promiscuous OMT for accepting different substrates, although the conversion is more effective for the first substrate (caffeic acid). On the other hand, CCoAOMT (Caffeoyl-CoA OMT that belong to the PL-OMTI group) catalyzes the methylation of Caffeoyl-CoA and 5-Hydroxy-Feruloyl CoA with the use of Mg^{+2} as a cofactor (figure 6, green pathway) (Lam et al., 2007).

A considerable economic interest in the genetic manipulation of lignin levels and/or its composition is present in several agriculture disciplines. In animal nutrition, lignin is considered an antinutritional component of forages and regarding bioethanol and biogas production, it negatively affects the biological energy conversion processes (Frei, 2013). Hence, the OMT genes are targets of genetic manipulation, in order to reduce the lignin amounts in some cases. Studies with plant mutants where CCoAOMT and COMT were downregulated showed a marked decrease in lignin amounts (Guo et al., 2001; Maury et al., 1999). However, this inhibition also showed effects on plant growth and resistance to pathogens (Frei, 2013; Maury et al., 1999). This is an example of how OMTs are important to different physiological responses *in planta*. Knowing that these proteins, despite family similarities at sequence level, are very different in structure can lead us to answers of the evolutionary process of plants and also give us tools to genetically improve the plants for economical interest.

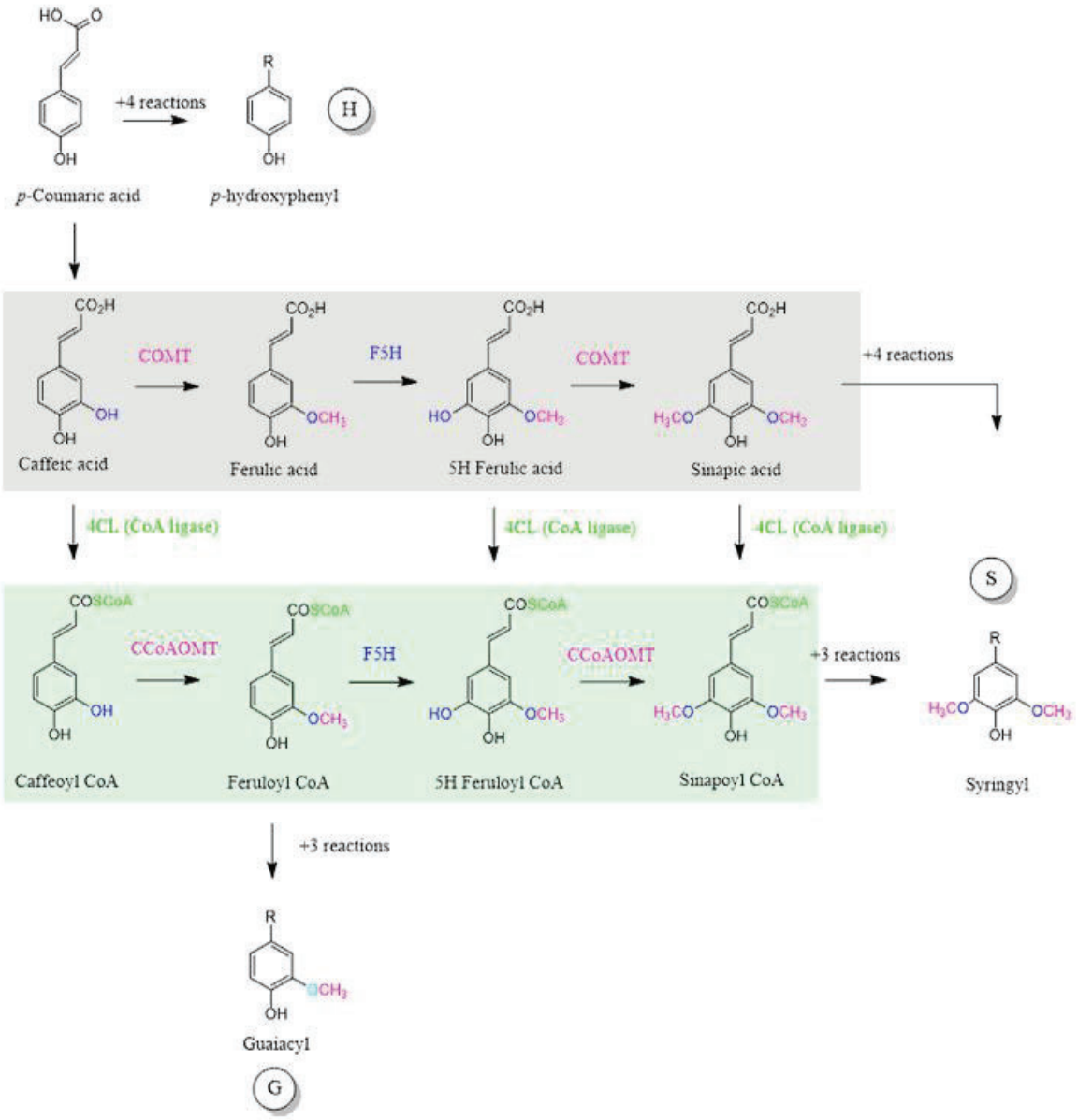


Figure 6: Biosynthesis of the three building blocks of lignin in a simplified model. The side group (R) of each monolignol could be a hydroxyl group or an ethyl aldehyde group, for example. Figure modified from Zhong et al. (2000).

2.3.2. Flavonoid OMTs

Flavonoids consist of a large family of molecules of at least 10,000 compounds registered to date. In plants, flavonoids are responsible for a big variety of biological activities including color, aroma of flowers and fruits, being highly involved in reproduction. In addition, flavonoids also help to protect plants from biotic and abiotic stresses, acting as a UV-filter but also displaying antimicrobial activity (Samanta et al., 2011). The biosynthesis of flavonoids starts on the primary

metabolism of phenylalanine and shares a common precursor with the phenylpropanoids synthesis, the cinnamic acid. Flavonoids are classified according to the carbonyl group present on the carbon 2, the presence or not of a double bond between carbon 2 and 3, or the presence of a hydroxyl group; resulting on the following categories: flavone, flavanone, flavonol, flavone, isoflavone, isoflavanone, and anthocyanin (Khalid et al., 2019).

The *O*-methylation of flavonoids has been proved to have a huge impact on its functionality. Overall, it is known that the methylation of a hydroxyl group of a flavonoid reduces their reactivity, increasing its microbial activity, shifting their role for a defensive action if needed (Ibrahim et al., 1998). In addition, the methylation of flavonoids have been shown to have a big impact also on its therapeutic activity. Studies showed that natural flavonoids that are usually prescribed as cancer chemoprotectants were highly more effective in their methylated form. The authors explain that the absence of a free hydroxyl group on these molecules avoids their rapid metabolism by the liver, dramatically reducing their depletion and increasing their bioavailability (Wen & Walle, 2006; Wilson et al., 2008).

Given the importance of the methylation for flavonoids, host organisms, such as *Escherichia coli*, have been effectively used to produce natural methylated flavonoids with the assistance of OMTs. A few examples are rhamnetin using the POMT7 from *Populus deltoides* (Sung et al., 2011), 3-*O*-methyl-kaempferol with the ROMT9 from *Oryza sativa* (Kim et al., 2010) and genkwanin with POMT7, also from *Populus deltoides* (Jeon et al., 2009).

2.3.3. Alkaloid OMTs

Alkaloids are a well known class of natural molecules produced by the secondary metabolism of plants. Because of its pronounced toxicity, they are used as a defense mechanism by the plants, against pathogens, insects or herbivores. To date, more than 12000 alkaloids have been identified in plants. These molecules were classified in four classes: monoterpene indole alkaloids, which resulted from the condensation of an amino acid (tryptophan) with a monoterpene derivative (secologanin) (Almagro et al., 2015); the benzylisoquinoline alkaloids, derived from tyrosine (Hagel & Facchini, 2013); the tropane alkaloids derived from the arginine and phenylalanine; and finally, the purine alkaloids, which are derived from purine (Ashihara et al., 2008). Because of its therapeutic effects, such as its anticancer effects, the alkaloids are of high interest especially for the pharmaceutical industry, being largely studied. Nowadays, with the increasing interest in the study of biosynthetic pathways looking forward to its transfer to

heterologous organisms, the knowledge of the reactions involved in alkaloids synthesis became a must for economic purposes.

The *O*-methyltransferases are highly involved in alkaloids synthesis from all kinds. The table 2 lists the representative alkaloid plant *O*-methyltransferases, their preferred substrates and their biological significance.

Table 2: Representative alkaloid plant *O*-methyltransferases and their biological significance.

Uniprot	Organism	Preferred	Biological	Reference
B0EXJ8	*Catharanthus roseus (16OMT)*	16-hydroxytabersonine	Vinblastine biosynthesis	(Levac et al., 2008)
I3V6A7	*Papaver somniferum (SOMT1)*	(S)-scoulerine	Noscapine biosynthesis	(Dang & Facchini, 2012)
I3PLQ6	*Papaver somniferum (SOMT2)*	(S)-scoulerine	Noscapine biosynthesis	(Dang & Facchini, 2012)
I3PLQ7	*Papaver somniferum (SOMT3)*	(S)-scoulerine	Noscapine biosynthesis	(Dang & Facchini, 2012)
Q6WUC2	*Papaver somniferum (PSOMT2)*	(S)-reticuline	Laudanine biosynthesis	(Ounaroon et al., 2003)
Q6WUC1	*Papaver somniferum (PSOMT1)*	(S)-norcoclaurine	(S)-Scoulerine	(Ounaroon et al., 2003)
Q7XB10	*Papaver somniferum (4'OMT2)*	3'-hydroxy-N-methylcoclaurine	Reticuline biosynthesis	(Ziegler et al., 2005)
Q7XB11	*Papaver somniferum (4'OMT1)*	3'-hydroxy-N-methylcoclaurine	Reticuline biosynthesis	(Pathak et al., 2013)
C7SDN9	*Papaver somniferum (N7OMT)*	norreticuline	Papaverine biosynthesis	(Pienkny et al., 2009)
Q8KZ94	*Lentzea aerocolonigenes*	4'-demethylrebeccamy	Rebeccamycin biosynthesis	(Singh et al., 2008)
Q5C9L7	*Thalictrum flavum subsp. Glaucum (6OMT)*	(S)-norcoclaurine	Coclaurine biosynthesis	(Robin et al., 2016)
Q8H9A8	*Coptis japonica*	tetrahydrocolumbamine	Palmatine biosynthesis	(Morishige et al., 2002)
Q9LEL6	*Coptis japonica (6OMT)*	norcoclaurine	Coclaurine biosynthesis	(Sato et al., 1994)
A0A077EWA5	Narcissussp.aff. pseudonarcissus	Norbelladine	Galanthamine biosynthesis	(Kilgore et al., 2014)

As previously mentioned, the methylation of a hydroxyl group usually leads to a reduction in its reactivity, altering the solubility of the molecule. On the case of alkaloids, it is not different, the methylation inverts the polarity of its electronegative moiety, changing the profile of the molecule and increasing its hydrophobicity and its steric hindrance, which brings important consequences for its biological activity (Wessjohann et al., 2014).

The benzylisoquinoline alkaloid morphine brings a good illustration of how the methylation alters the property of an alkaloid. The figure 7 shows the structure of the morphine, which possesses a methyl group attached to the nitrogen of its benzylisoquinoline ring. The removal of this methyl group results in the molecule normorphine, which leads to a 6-fold less *in vivo* analgesic activity due to the increase of the polarity of the molecule, preventing the passage through the blood-brain barrier (Karlsson et al., 1988). For codeine, illustrated by the (A) methylation on the hydroxyl group of the carbon 3 of morphine, the *in vitro* activity is reduced by 200-fold, showing that the methylation on the wrong position can also bring negative effects for the therapeutic effects of the molecule. In this case, the negative effect of the methylation is caused by the absence of availability of the hydrogen-bond which is crucial for the interaction of morphine with its receptor. However, codeine is still a widely used drug, the reason is that *in vivo*, the hepatic enzyme CYPD6 (Cytochrome P450 2D6) performs a demethylation of codeine, converting a portion of the administered codeine into morphine. Finally, the heterocodeine which is represented by the (B) methylation of the hydroxyl group of the carbon 6 of morphine, is produced semi-synthetically from the morphine, and presents 2-fold more activity *in vivo* than morphine. The potency of this molecule is due to the same reason that led to the normorphine to lose its therapeutic properties, however on this case, the additional methyl increases the lipophilicity of the drug, facilitating its passage into the central nervous system (Barreiro et al., 2011; DeBono et al., 2015). Due to the high importance of alkaloids for medicine, Chapter III will describe in more details their use and the limitations associated with the production of these molecules.

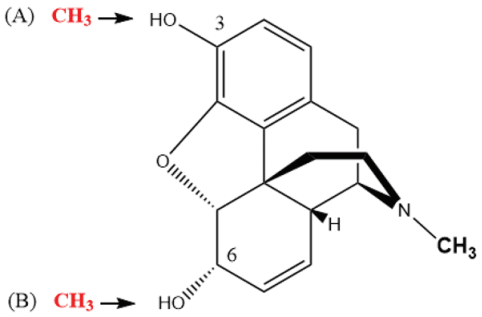


Figure 7: Chemical structure of the natural product morphine. (A) Methylation of carbon 3 leading to codeine, or 3-methylmorphine; (B) Synthetic methylation on carbon 6, leading to heterocodeine or 6-methoxymorphine.

2.4. Evolution of new functionalities in OMTs

Biological evolution cannot be explained as an engineering process. Evolution is limited to the organic material availability and to the biochemical tools of the system. New functions arise from the modification of genetic elements, and it is in this manner that novel enzymes, pathways and metabolism evolve, by different random assemblies. Since evolution follows no plan, it may generate neutral, negative or advantageous properties to an organism (Jacob, 1977).

The figure 8 shows a phylogenetic tree of representative PL-OMTII group members. The resulting tree shows that the selected OMTs segregate into 2 major clusters, the first consisting of mainly phenylpropanoid OMTs and the other consisting of diverse OMTs. The phenylpropanoid OMTs are believed to be mostly involved in the methylation of caffeic acid, a reaction that occurs in primitive forms of land plants, whereas diverse OMTs from the second cluster are thought to have evolved later in response to a more diverse set of evolutionary forces (Lam et al., 2007; Morris & Facchini, 2019).

The phenylpropanoids cluster is subdivided into 2 subgroups. The biggest subgroup contains mainly caffeic acid OMTs and promiscuous flavone OMTs, which may have evolved from the duplication and subsequent functional divergence of the ancestral gene. Interestingly, the second subgroup is represented by a sole protein scoulerine 9-OMT (CjSOMT). Scoulerine is a benzylisoquinoline alkaloid and surprisingly this OMTs is not clustered together with other benzylisoquinoline alkaloid OMTs (yellow subgroup). According to Morris and Facchini (2019), the scoulerine OMTs may have had a monophyletic origin and on the case of CjSOMT, its promiscuity for other substrates, such as columbamine may be interpreted as maintenance of an ancestral enzyme feature, which corresponds to its particular location on the tree (Morishige et al., 2002; Morris & Facchini, 2019).

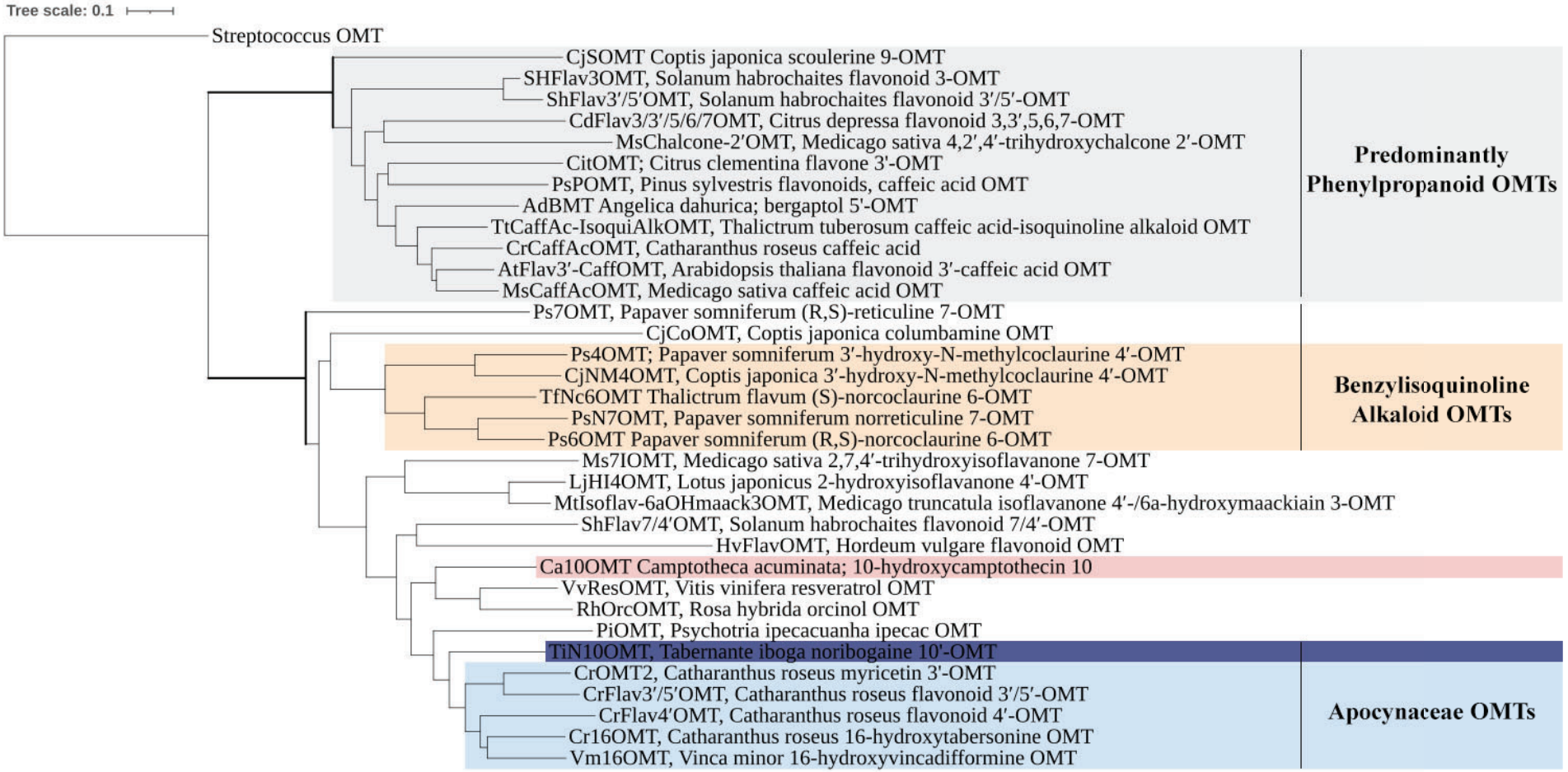


Figure 8: Phylogenetic relationships among representative functionally characterized plant O-methyltransferases.

Figure 8 (continuation): The phylogenetic tree was generated using the neighbor-joining method available at the Cobalt tool (https://www.ncbi.nlm.nih.gov/). The tree design was performed with the iTOL tool (https://itol.embl.de/). Representative plant OMT sequences and clade assignments were adapted from Salim et al. (2018). The plant species and accession numbers are: CjSOMT: *Coptis japonica* scoulerine 9-OMT (BAA06192.1); ShFlav3′/5′OMT: *Solanum habrochaites* flavonoid 3′/5′-OMT (ADZ76433.1); SHFlav3OMT: *Solanum habrochaites* flavonoid 3-OMT (AGK26768.1); MsChalcone-2′OMT: *Medicago sativa* 4,2′,4′-trihydroxychalcone 2′-OMT (P93324.1); CdFlav3/3′/5/6/7OMT: *Citrus depressa* flavonoid 3,3′,5,6,7-OMT (BAU51794.1); AdBMT: *Angelica dahurica* bergaptol 5'-OMT (AEO21927.1); TtCaffAc-IsoquiAlkOMT: *Thalictrum tuberosum* caffeic acid-isoquinoline alkaloid OMT (AAD29841.1); CrCaffAcOMT: *Catharanthus roseus* caffeic acid (AAK20170.1); MsCaffAcOMT: *Medicago sativa* caffeic acid OMT (AAB46623.1); AtFlav3′-CaffOMT: *Arabidopsis thaliana* flavonoid 3′-caffeic acid OMT (AAM10127.1); PsPOMT: *Pinus sylvestris* flavonoids caffeic acid OMT (AQX17825.1); CitOMT: *Citrus clementina* flavone 3'-OMT (XP 006427989.2); Ps7OMT: *Papaver somniferum* R,S-reticuline 7-OMT (AAQ01668.1); CjCoOMT: *Coptis japonica* columbamine OMT (Q8H9A8.1); CjNM4OMT: *Coptis japonica* 3′-hydroxy-N-methylcoclaurine 4′-OMT (Q9LEL5.1); Ps4OMT: *Papaver somniferum* 3′-hydroxy-N-methylcoclaurine 4′-OMT (Q7XB10.1); TfNc6OMT: *Thalictrum flavum* S-norcoclaurine 6-OMT (AAU20765.1); Ps6OMT *Papaver somniferum* R,S-norcoclaurine 6-OMT (AAQ01669.1); PsN7OMT: *Papaver somniferum* norreticuline 7-OMT (ACN88562.1); Ms7IOMT: *Medicago sativa* 2,7,4′-trihydroxyisoflavanone 7-OMT (AAC49927.1); MtIsoflav-6aOHmaack3OMT: *Medicago truncatula* isoflavanone 4′-/6a-hydroxymaackiain 3-OMT (AAY18581.1); LjHI4OMT: *Lotus japonicus* 2-hydroxyisoflavanone 4'-OMT (BAC58013.1); HvFlavOMT: *Hordeum vulgare* flavonoid OMT (CAA54616.1); ShFlav7/4′OMT: *Solanum habrochaites* flavonoid 7/4′-OMT (ADZ76434.1); Ca10OMT: *Camptotheca acuminata* 10-hydroxycamptothecin OMT (AWH62806.1); RhOrcOMT: *Rosa hybrida* orcinol OMT (CAH05087.1); VvResOMT: *Vitis vinifera* resveratrol OMT (CAQ76879.1); PiOMT: *Psychotria ipecacuanha* ipecac OMT (BAI79243.1); TiN10OMT: *Tabernante iboga* noribogaine 10'-OMT (AXF35975.1); CrFlav3′/5′OMT: *Catharanthus roseus* flavonoid 3′/5′-OMT (AAM97498.1); CrOMT2: *Catharanthus roseus* myricetin 3'-OMT (Q8GSN1.1); CrFlav4′OMT: *Catharanthus roseus* flavonoid 4′-OMT (AAR02419.1); Vm16OMT: *Vinca minor* tabersonine/vincadifformine 16′-OMT (QBY35563.1); Cr16OMT: *Catharanthus roseus* 16-hydroxytabersonine OMT (ABR20103.1). The Streptococcus OMT (WP_191870180.1) was used to root the tree.

The second cluster contains the diverse OMTs which are mostly subgrouped regarding their substrate preference. The yellow subgroup contains benzylisoquinoline alkaloid OMTs. The existence of OMTs with the same substrate preferences from different species clustering together, such as Ps4OMT from *P. somniferum* and CjNM4OMT from *C. japonica*, strongly support the gene conservation across species.

Another subgroup can be observed highlighted in blue (Apocynaceae OMTs). Interestingly, these OMTs are subgrouped together even if their preferred substrates are chemically different, such as the flavonoid OMT (CrFlav4'OMT) and the monoterpene indole alkaloid OMT (Cr16OMT). Since both enzymes share the same ancestral gene, it is possible that after its duplication, an alkaloid specific OMT brought evolutionary advantage within the *C. roseus* specialized metabolism. The Cr16OMT also cluster with the tabersonine/vincadifformine 16-O-methyltransferase (Vm16OMT), both proteins are structurally close and share common substrates, however Vm16OMT can methylate 16-hydroxylated derivatives of tabersonine, vincadifformine and lochnericine, whereas Cr16OMT is specific to 16-hydroxytabersonine. According to (Lam et al., 2007), structurally similar proteins that act on a variety of different substrates indicates a divergent evolution, which is commonly defined as what occurs when two groups of the same species evolve different traits within those groups in order to accommodate for differing environmental pressures.

The Ca10OMT, shaded in pink, is a special case of a flavonoids/benzylisoquinoline alkaloids promiscuous enzyme, and its position on the tree suggests that its ancestor gene was also a protein which accepted alkaloids, flavonoids and phenols. This hypothesis is supported by the subgroup shaded in blue, where monoterpene indole alkaloid (MIA) and flavonoid OMTs from *C. roseus* are clustered. The ancestor of this clade also gave origin to the noribogaine-10-O-methyltransferase (TiN10OMT), shaded in purple, an important kind of MIA, classified as iboga type MIA.

Overall, alkaloid OMTs occupies various subgroups on the phylogenetic tree, showing convergent or repeated evolution, in which the C-terminal of these proteins are highly conserved for SAM binding, whereas the putative alkaloid binding site is located on the N-terminus and are more divergent, reflecting their substrate specificity (Morishige et al., 2002; Lam et al. 2007).

2.5. Conclusion

This review focused on giving an overall vision on the high diversity of OMTs and their importance, therefore, a complete examination of known plant OMTs is outside the goal of this chapter. A complete understanding of *O*-methyltransferases function and evolution of specific enzymes is of extreme importance for the understanding of the origin of biosynthetic pathways of primary or specialized metabolites and will contribute to reach a full comprehension of natural products synthesis.

3. Medicinal alkaloids

Since very early in human history, the chemical diversity of plants acted as beneficial promoters for human life (Hartmann, 2007). Nowadays, the active molecules from these plant extracts have been extensively studied for their biosynthesis, in order to achieve the industrial and rational use of their biological activities, especially by pharmaceutical companies. Known for their activity that ranges from medicinal to toxic, alkaloids require a controlled dosage and moreover, production optimization due to the low quantities accumulated in the plant sources (Bhambhani et al., 2021; Dey et al., 2020). Plants from several families are known to produce medicinal alkaloids such as Cornaceaes, Apocynaceaes, Papaveraceaes, Berberidaceaes, among others. Despite possessing common ancestors, these plants possess unique alkaloids with very particular biological properties.

On table 3, representative natural medicinal alkaloids are presented from different species. Despite in some cases, the alkaloids possess the same medicinal properties, their chemical structures can highly differ and their mode of action will vary accordingly. The four main classes of alkaloids are presented. The most important monoterpene indole alkaloids are represented by vincristine and vinblastine. The morphine and codeine represent the benzyliquinoline alkaloids. The controversial cocaine represents the tropane alkaloid and the incredible quinine the purine alkaloids.

The plant culture and harvesting are still important steps in the process of obtaining natural molecules of pharmacological interest. Hence, many plants have been identified and exploited for their ability to provide these compounds which, in practical terms, can be translated as an extremely high number of lands dedicated to this purpose. The cost-benefice of this practice does not add up. The amount in which these molecules are found in the plants is incredibly low and tones of leaves are needed in order to produce a few grams of most of the alkaloids listed on the table 3. *C. roseus* vinblastine and vincristine, for example, are found at the order of 2 ppm in leaves, meaning that 1 ton of leaves is necessary for the extraction of 1 g of vinblastine, making this molecule extremely expensive (Courdavault et al., 2020). This continuous exploitation resulted in years of costly and laborious medicine production. Despite being a technical problem for a long time, the need to change the way these molecules are produced became urgent in the face of an environmental crisis. The use of plant sources needs to find an end, both regarding nature preservation but also due to the strong dependence of the health system upon these sources.

Hence, an increasing interest regarding the biosynthesis of these molecules came up in the last years. The fast advancements in transcriptomic and metabolomics accelerated the process of mastering the biosynthesis of these molecules.

Table 3: Representative natural medicinal alkaloids.

Alkaloid	Biological activities	Plant source	References
Vincristine **Vinblastine**	Antileukemic, antilymphoma, antineuroblastoma, and antisarcoma	*Catharanthus roseus*	(Alam et al., 2017)
Camptothecin	Antitopoisomerase and anti-HIV	*Camptotheca acuminata, Nyssaceae*	(Li et al., 2009)
Colchicine	Anti-gout, anti-inflammation, and treats familial Mediterranean fever	*Colchicum and Gloriosa species*	(Nett et al., 2020)
Reserpine	Anti-hypertensive and anti-psychotic	*Rauwolfia species*	(Kumar et al., 2016; Lobay, 2015)
Ajmalicine	Anti-hypertensive		
Ajmaline			
Morphine	Analgesic and central nervous system stimulant	*Papaver somniferum*	(Carlin et al., 2020; Stermitz, 2000)
Codeine	Analgesic, antidiarrheal, and antitussive		
Papaverine	Vasodilatory and antispasmodic		
Berberine	Antimicrobial, antitumor, anti-hyperglycemic, antimalarial, and anti-inflammation; Alzheimer's disease treatment	*Berberis species*	(Hou et al., 2020)
Caffeine	Autonomous nervous system stimulant, anti-inflammation; improves cognitive performance	*Cocoa beans, guayusa, yaupon holly*	(Bhambhani et al., 2021)
Nicotine	Treatment of nicotine dependence	*Nicotiana species*	(Sajja et al., 2016)
Quinine	Antimalarial	*Cinchona succirubra*	(Achan et al., 2011)
Cocaine	Stimulant	*Erythroxylum coca*	(Grynkiewicz & Gadzikowska, 2008)

The *C. roseus* plants became a model plant on this matter due to the extensive studies of its metabolism and the unrevealing of all the steps that leads to the biosynthesis of monoterpene indole alkaloids. *C. roseus* is endemic to Madagascar and was discovered by Carl von Linné in 1759, who called it the *Vinca rosea*, placing this plant under the Genus Vinca, however this was

changes on century latter by George Don, who placed the Madagascar periwinkle under the Genus Catharanthus, which counts with 8 species nowadays (Van der Heijden et al., 2004).

The *C. roseus* dimeric alkaloids vinblastine and vincristine are considered as one of the most important plant alkaloids for medicine. The anticancer properties of these alkaloids rely on their capacity to avoid microtubules polymerization, by binding to tubulins and avoiding their assembly, which, consequently, stops the cell division in metaphasis, inducing cell apoptosis (Gigant et al., 2005). Both MIAs are used in chemotherapy treatments: the vinblastine is used against Hodgkin's diseases, while the vincristine is used against lymphoblastic leukaemia. Logically, due to the toxicity of these molecules, a prolonged treatment should be avoided regarding the long list of undesirable secondary effects.

Due to their powerful activity and efficiency, researchers have been developing new non-natural dimeric alkaloids that are less toxic than the natural one. The well-known vindesine and vinorelbine are examples of therapeutic molecules produced by semi-synthesis, which opens a new and promising field for the evolution of the use of plant alkaloids in human health (Jones & Burris, 1996).

Other medicinal alkaloids are present in *C. roseus*. Various alkaloids were observed to have hypoglycemic activity such as vindoline, tetrahydroalstonine, catharanthine and lochericine (Marles & Farnsworth, 1995). In addition, ajmalicine and serpentine are used as anti-hypertensive and anti-neuro-inflammatory agents and yohimbine is mainly used in treatments for erectile dysfunction (Almagro et al., 2015).

4. *Catharanthus roseus* and the monoterpene indole alkaloid biosynthesis

4.1. *C. roseus* MIAs

Catharanthus alkaloids, known previously as the *Vinca* alkaloids consist of around 130 alkaloids, mostly monoterpene indole alkaloids (MIAs). Strictosidine is the central intermediate which undergoes a series of reactions to produce over 3,000 known MIAs (Brown et al., 2015). The MIAs can be divided into two classes, the monomeric and the dimeric MIAs. The monomeric alkaloids of *C. roseus* can be represented by ajmalicine, catharanthine or vindoline. Conversely, the dimeric alkaloids are represented by the vinblastine and vincristine (figure 9).

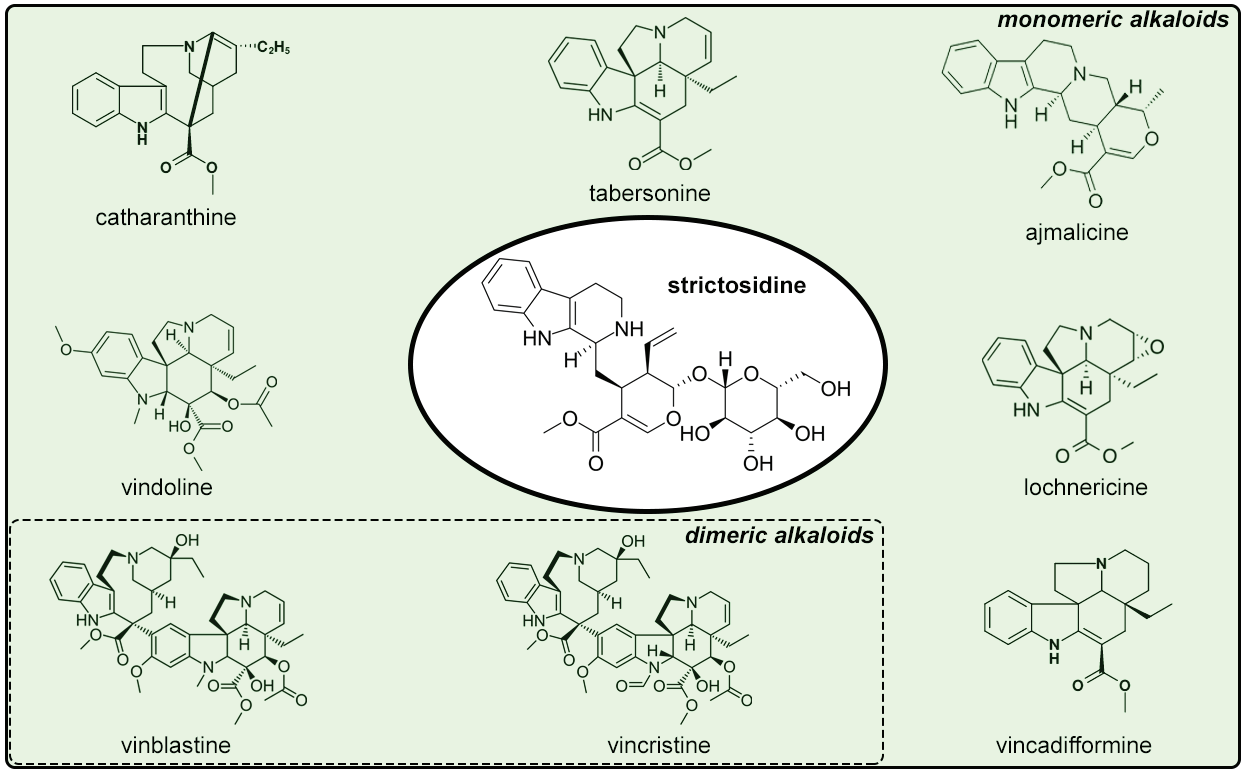


Figure 9: *C. roseus* monoterpene indole alkaloids. The monomeric alkaloids are represented by catharanthine, tabersonine, ajmalicine, vindoline, strictosidine (the common precursor), lochnericine and vincadifformine. The dimeric alkaloids are represented by vincristine and vinblastine.

Since 2010, *C. roseus* has become a model plant regarding MIA metabolism due to the advancements on the elucidations of its highly complex pathway, in which more than 30 enzymatic steps are involved. This advances were possible because of the numerous projects in transcriptomics analysis of this plant, such as the Phytometasyn (http://www.phytometasyn.ca) (Xiao et al., 2013), Medicinal Plant Genomics Consortium (http://medicinalplantgenomics.msu.edu/) (Góngora-Castillo et al., 2012) and the not concluded project 1000 Plants Initiative (www.on-ekp.com/) that offer the possibility to obtain simultaneously qualitative (raw sequence) and quantitative (mRNA abundance estimation) data (Dugé de Bernonville et al., 2015). Hence, the full characterization of important MIA pathways such as vindoline and catharanthine pathways has been accomplished as well as their global tissue architecture.

The MIA pathway can be divided in three different blocs: 1) The tryptamine synthesis from the shikimate pathway; 2) the synthesis of the monoterpenic precursor secologanin originating from the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway; 3) the synthesis of strictosidine and the following pathways that lead to catharanthine and vindoline, that after condensation, produce the dimeric MIAs vinblastine and vincristine (figure 10). It is important to notice that the division presented in the figure 10 is not directly related to pathway architecture. The *C. roseus* MIA pathway architecture is extremely complex because it takes in consideration both tissue and intra cellular organization. The full description of the first 2 blocs would make this literature review extremely extensive, thus we will be focusing on the third bloc, especially on the production of vindoline from tabersonine and the full explanation of the involved mechanisms.

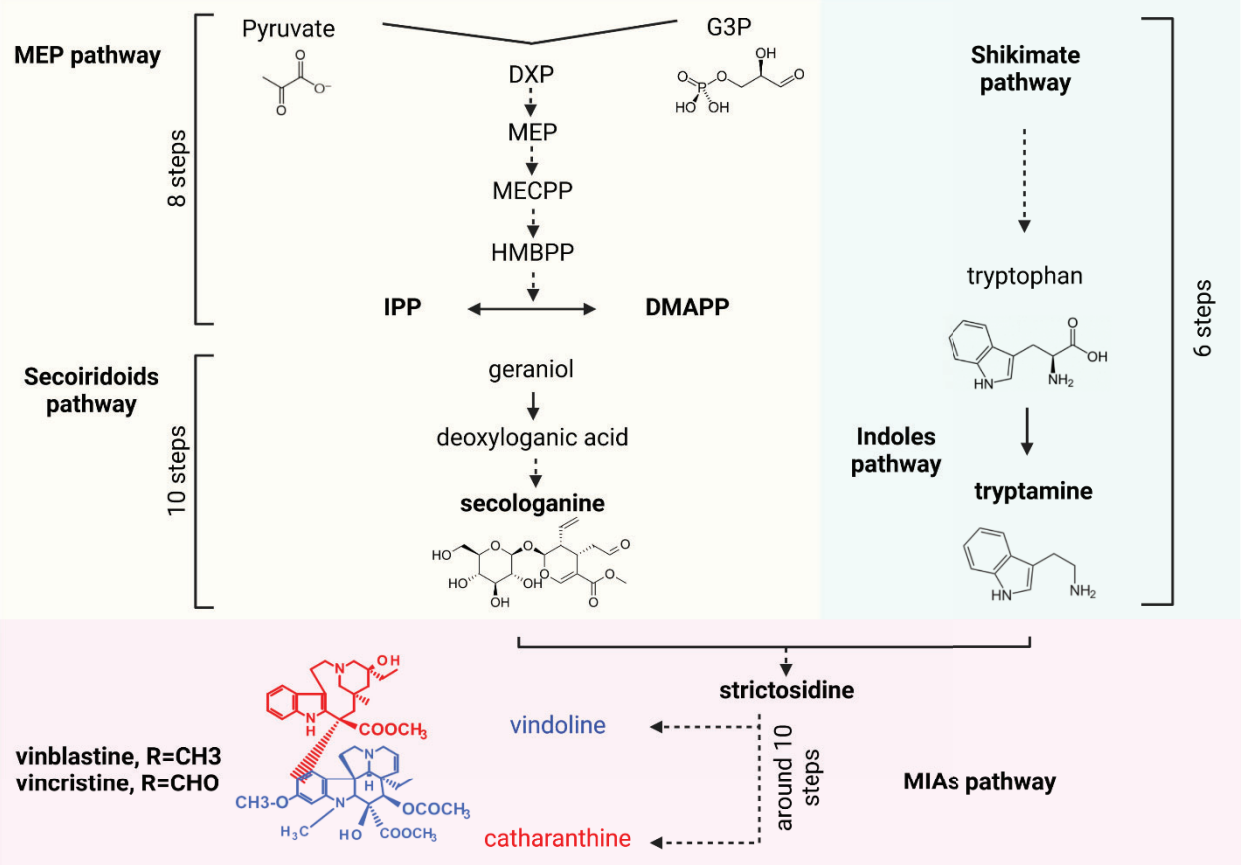


Figure 10: A summarized view of MIA biosynthesis in *C. roseus*. DXP: 1-deoxy-D-xylulose 5-phosphate; **MEP**: 2-C-methyl-D-erythritol-4-phosphate; **MECPP**: 2-C-methyl-D-erythritol-2,4-cyclodiphosphate; **HMBPP**: 4-hydroxy-3-methylbut-2-enyl diphosphate; **IPP**: isopentenyl diphosphate; **DMAPP**: dimethylallyl diphosphate. A solid arrow stands for one enzymatic step, while a dotted line represents several enzymatic steps.

4.2. *C. roseus* MIA biosynthesis

The medicinal alkaloids produced by *C. roseus* are mostly present in its leaves, where their assembly is a complex organization between mainly four types of cells: the internal phloem associated parenchyma (IPAP), epidermis, laticifer and idioblast cells. In each of these cells, several subcellular compartments are also involved, such as nucleus, vacuole, endoplasmic reticulum and cytosol (Courdavault et al., 2014). Recent techniques such as single-cell UPLC-MS/MS were required to establish cellular distribution of alkaloids in *C. roseus* leaves. The protein subcellular localization was explored by biolistic-mediated transient transformation of *C. roseus* cells, followed by fluorescence microscopy visualization.

The IPAP cells will be the first type of cells to take part on the pathway, hosting the 16 first steps, including the whole MEP pathway, followed by 8 reactions from the secoiridoid pathway, which gives origin to the loganic acid (Courdavault et al., 2014) (figure 10).

The second group of cells are the epidermis cells, where the main reaction of the MIA pathway takes place. From IPAP cells, loganic acid is translocated to epidermal cells, where it will be transformed into secologanin by two enzymatic reactions. The translocation of loganic acid to epidemis cells was recently elucidated with the identification of three transporters called CrNPF2.4, CrNPF2.5 and CrNPF2.6 (Larsen et al., 2017). In the vacuole of epidermis cells, tryptamine and secologanin are coupled by the enzyme strictosidine synthase (STR), representing the first enzymatic steps of the *C. roseus* MIA synthetic pathway. The strictosidine will be transported to the nucleus, where it will find the strictosidine-β-D-glucosidase, responsible for the hydrolysis of the glucose moiety from strictosidine, generating the unstable aglycon, which is rapidly converted to an reactive dialdehyde intermediate, a common precursor for the vast majority of MIAs (Guirimand et al., 2010) (figure 10, figure 11). The proposed hypothesis of the mechanism in which strictosidine is translocated to the nucleus will be presented in the section "4.3: Role of MIAs in *C. roseus* metabolism".

Following the formation of 4,21-dehydrogeissoschizine from strictosidine, eight enzymatic steps take place in the cytosol of epidermis cells, to produce catharanthine or tabersonine (figure 11). Vindoline is produced after tabersonine is hydroxylated by tabersonine-16-hydroxylase (T16H2), methylated by tabersonine-16-O-methyltransferase (16OMT), epoxidated by tabersonine 3-oxygenase (T3O), reduced by tabersonine 3-reductase (T3R) and *N*-methylated by 3-hydroxy-16-methoxy-2,3-dihydrotabersonine N-methyltransferase (NMT), generating the product desacetoxyvindoline in the cytosol of epidermis cells. It is worth mentioning that NMT localization was very recently characterized in peroxisomes, pointing out a new layer of complexity to the subcellular organization of the MIA biosynthetic pathway (Koudounas et al., 2021).

The last two groups of cells are the specialized laticifer and idioblast cells which host the final two steps of vindoline biosynthesis (Guirimand et al., 2011). The desacetoxyvindoline is transferred from epidermis cells to the laticifer or idioblast cells, hydroxylated by the cytosolic desacetoxyvindoline 4-hydroxylase (D4H) and *O*-methylated by the cytosolic deacetylvindoline 4-O-acetyltransferase (DAT). These two last steps will yield vindoline that will be stored in the vacuole, considered as the most abundant alkaloid in *C. roseus* leaves (Carqueijeiro et al., 2016;

Yamamoto et al., 2019). Another alkaloid that is also produced in these cells, is the vindorosine, a low-accumulated MIA that is also produced from tabersonine. Prior to the explanation of the mechanisms involved in vindorosine production, two concepts need to be presented, due to their extremely importance to the *C. roseus* MIA biosynthesis: protein isoforms and enzyme promiscuity.

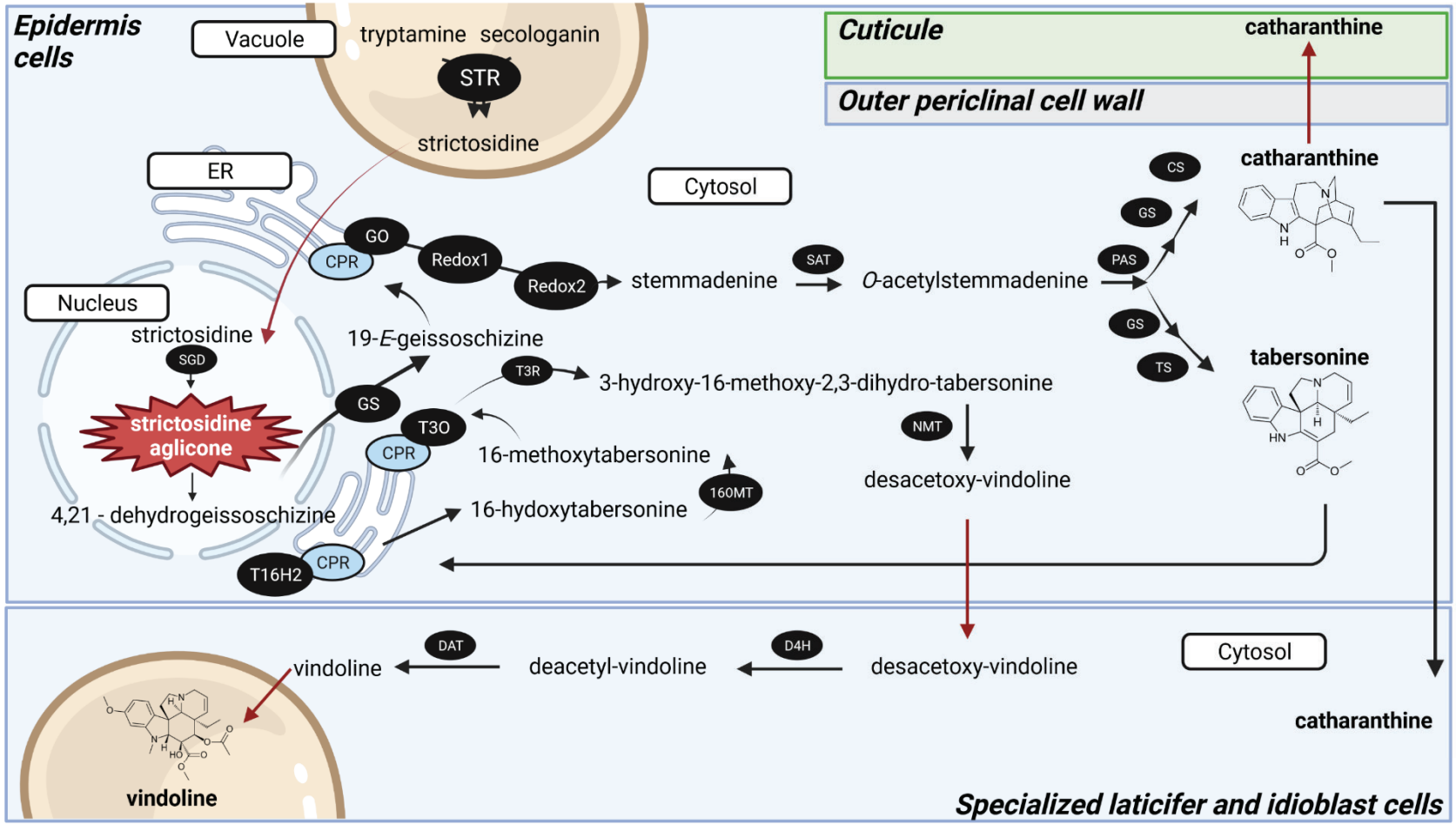


Figure 11: *C. roseus* MIA biosynthetic pathway: An integrated model of its cellular and subcellular architecture. STR: strictosidine synthase; **SGD**: strictosidine β-glucosidase; **GS**: geissoschizine synthase; **GO**: geissoschizine oxydase; **REDOX1**: cinnamyl alcohol dehydrogenase-like enzyme; **REDOX2**: aldoketo–type reductase; **SAT**: stemmadenine-O-acetyltransferase; **PAS**: precondylocarpine acetate synthase; **CS**: catharanthine synthase; **TS**: tabersonine synthase; **T16H2**: tabersonine 16-hydroxylase 2; **16OMT**: 16-hydroxytabersonine O-methyltransferase; **T3O**: tabersonine 3-oxygenase; **T3R**: tabersonine 3-reductase; **NMT**: 2,3-dihydro-16-methoxytabersonine N-methyltransferase; **D4H**: desacetoxyvindoline 4-hydroxylase; **DAT**: deacetylvindoline 4-O-acetyltransferase. Adapted from Courdavault et al. (2014)

4.2.1. Enzyme promiscuity: T3O, T3R, NMT, DAT and D4H

In general, enzymes are related to the word “specificity”, in which a perfect binding between the substrate and the enzyme will occur, leading to the famous “Lock-and-key” model. It is true that many enzymes are indeed specific for only one substrate, however, many other enzymes are not so specific and are able to act in other substrates besides their primary ones, producing other products or even catalyzing different reactions. This ability is referred to as “promiscuity”.

Enzyme promiscuity is a broad subject and can have different meanings in different contexts. In evolutionary biology, a promiscuous enzyme is an enzyme in which its functions are not related to any physiological relevance, meaning that its biological function remains unknown (Copley, 2015). However, in biochemistry, this term is used to describe an enzyme possessing catalytic active sites that allow different chemical transformations, involving different functional groups and/or catalytic mechanisms, and does not take in consideration which is its physiological relevance (Bornscheuer & Kazlauskas, 2004).

Several kinds of promiscuity were described in nature. Product promiscuity is the less studied mechanism and refers to the ability of an enzyme to produce different products from a single substrate, for example, the aureusidin synthase, from snapdragon, that is able to catalyze the hydroxylation and/or oxidative cyclization of the precursor 2′,4′,6′,3,4-pentahydroxychalcone (Nakayama, 2000). Differently, catalytic promiscuity is the ability of an enzyme to perform different reactions by using different substrates, as the 2-hydroxyisoflavanone dehydratase of soybean, which is able to perform the dehydration of 2- hydroxyisoflavanone and the esterase activity when in contact with carboxylic esters, as *p*-nitrophenyl butyrate (Akashi et al., 2005). Lastly, substrate promiscuity is the most studied mechanism of enzyme promiscuity, which is the ability of an enzyme to use different substrates in the same type of reaction. This mechanism can be found in a diverse range of plants and metabolic pathways, including the vindoline pathway.

On the figure 12, a scheme of the fate of tabersonine has been drawn. In *C. roseus* leaves, tabersonine can be transformed into two final products, vindorosine, that is a product of no pharmaceutical interest, or the highly desired vindoline. At first glance, it is possible to recognize that several enzymes are responsible for more than one reaction, such as T3O and T3R (marked in red in figure 12).

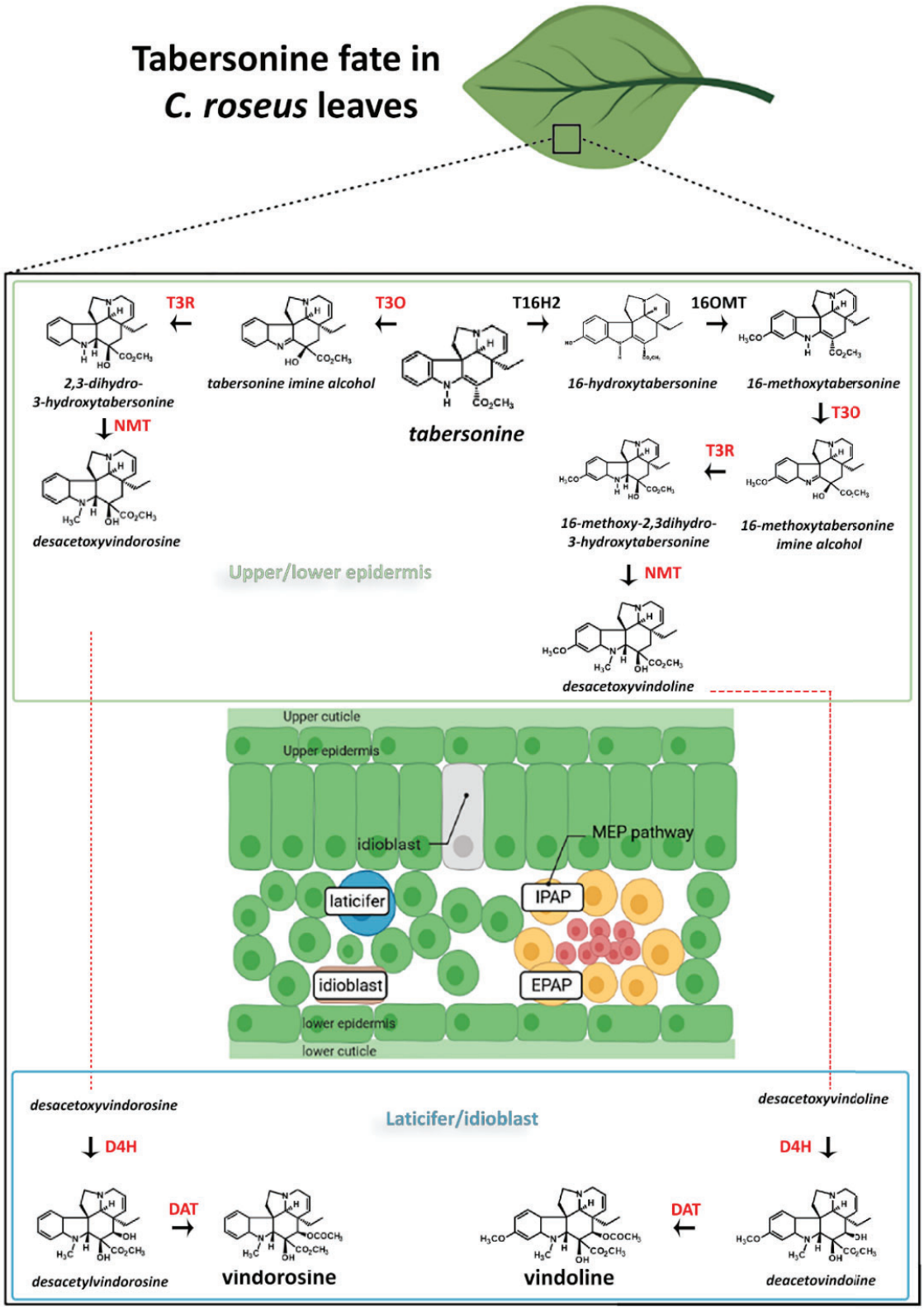


Figure 12: Tabersonine fate in *C. roseus* leaves and its tissue distribution. **T16H2**: tabersonine 16-hydroxylase 2; **16OMT**: 16-hydroxytabersonine O-methyltransferase; **T3O**: tabersonine 3-oxygenase; **T3R**: tabersonine 3-reductase; **NMT**: 2,3-dihydro-16-methoxytabersonine N-methyltransferase; **D4H**: desacetoxyvindoline 4-hydroxylase; **DAT**: deacetylvindoline 4-O-acetyltransferase. The enzymes marked in red are promiscuous enzymes.

It has been previously reported that tabersonine 16-hydroxylase (T16H2) is likely to play the key role in controlling the direction of both vindoline and vindorosine pathway, where VIGS analysis of T16H2 led to the accumulation of vindorosine precursors (Besseau et al., 2013; Yamamoto et al., 2019). In addition, RNA-seq data showed an increased gene expression of T16H2 in leaf tissue compared to steam tissue, where considerably lower amounts of vindoline are found (Yamamoto et al., 2019).

4.2.2. Protein isoforms

Proteins that possess a high identity percentage between their sequences and perform similar functions are called isoforms. These isoforms may present different kinetic properties and within the plant, can be expressed in a tissue-specific manner and possess different regulatory configurations. Regarding their origin, they may originate from two divergent mechanisms. First, a single gene can encode two or even more isoforms by a process called alternative splicing that can be briefly described by a process in which a pre-mRNA is processed and spliced post-transcriptionally to create more than one translatable mRNAs. Or else, the second mechanism is simply the existence of two or more closely related genes in which, each one will code for one different isoform (Gunning & Hardeman, 2018).

In the last section, the important role of T16H2 in the regulation of the tabersonine's fate was discussed. This protein is found in two isoforms in *C. roseus*, where the first isoform T16H1 was described by St-Pierre & De Luca (1995) and it was believed that to be the key protein on the regulation of vindoline in *C. roseus*, what changed in 2013, when Besseau and collaborators found a second isoform T16H2 that showed much more impact in vindoline biosynthesis (Besseau et al., 2013). It was found that despite both of the proteins displaying high affinity for tabersonine and possessing similar kinetic properties, they differ dramatically in transcript distribution *in planta*. The first isoform T16H1 was found to be expressed mainly in flowers, while T16H2 presented the same expression profile that other MIA genes – highly expressed in young leaves, more specifically, in epidermis cells. Finally, T16H2 VIGS experiments showed a reduction of vindoline, being the crucial point to determine that this isoform is indeed the key point of vindoline synthesis.

4.3. Role of MIAs in *C. roseus* metabolism

The MIA synthesis in *C. roseus* can be observed starting from 2 to 3 weeks after the germination and their concentration changes regarding the plant life cycle. Their distribution is also not homogenous. The higher MIA concentrations are found in developing leaves and decrease when they reach maturity. The type of MIA will also play a role in its distribution, for example, the dimeric alkaloids are present exclusively in the aerial parts of the plant, as well as vindoline. Conversely, serpentine and ajmalicine will be found mainly in the roots. Notably, the concentration differs from one MIA to another. Final dimeric alkaloids, vincristine and vinblastine will be found at concentration of around 1 and 4 ppm, respectively, while vindoline and catharanthine, the precursors of the dimeric MIAs, will be found at the concentration of 400 and 200 ppm, respectively (Guirimand et al., 2011; Yamamoto et al., 2019).

As previously discussed, SMs can display several functions in plants regarding their needs and the current biotic and abiotic stresses they are overcoming. In the case of MIAs, it is proposed that they are responsible for the plant defense against herbivores. Considering the high toxicity of MIAs, the strategy used by plants to fight the biotic attack avoiding its own intoxication is remarkable. One of the strategies relies on the complexity of both tissue organization and subcellular localization of *C. roseus* MIAs.

Strictosidine is an alkaloid produced in the vacuole of epidermis cells, however, the enzyme which is responsible for the deglucosylation of strictosidine, strictosidine-β-D-glucosidase (SGD), is located in the nucleus. Under physiological conditions, the NFP2.9 transporter allows externalization of limited amounts of strictosidine out of the vacuole followed by a passive diffusion into the nucleus to allow downstream MIA synthesis (Payne et al., 2017). Upon leaf attacks, the situation is dramatically different. An elegant hypothesis was indeed proposed by Guirimand et al. (2010), that states that the substrate and the enzyme will be in contact only upon cell damage. Briefly, during an herbivore attack, both vacuolar and nuclear membranes will be disrupted, leading to the substrate-enzyme interaction. Consequently, this overproduction of the strictosidine aglycon will not be metabolized by the MIA biosynthetic pathway, which will lead to a lethal massive protein cross-linking and precipitation of strictosidine aglycon, preventing the herbivores to continue fasting the leaves (figure 13). In order to avoid any kind of leakage of SGD proteins from the nucleus, SGD assembles into aggregate through multimerization, which reinforces even more the compartmentalization of this pathway (Guirimand et al., 2010).

A recent study showed that SGD overcomes an alternative splicing, which is a process that leads to the formation of a shorter isoform of this enzyme, which was called short SGD (shSGD) by the authors. This isoform shows no β-glucosidase activity; however, it can interact with SGD and causes the disruption of SGD multimers, thus inhibiting SGD activity and impacting downstream MIA synthesis. Interestingly, shSGD disrupts the metabolic channeling of downstream biosynthetic steps by blocking the tetrahydroalstonine synthase activity in the nucleus. The shSGD is an example of another kind of regulator mechanism of MIA biosynthesis for defense purposes (Carqueijeiro et al., 2021).

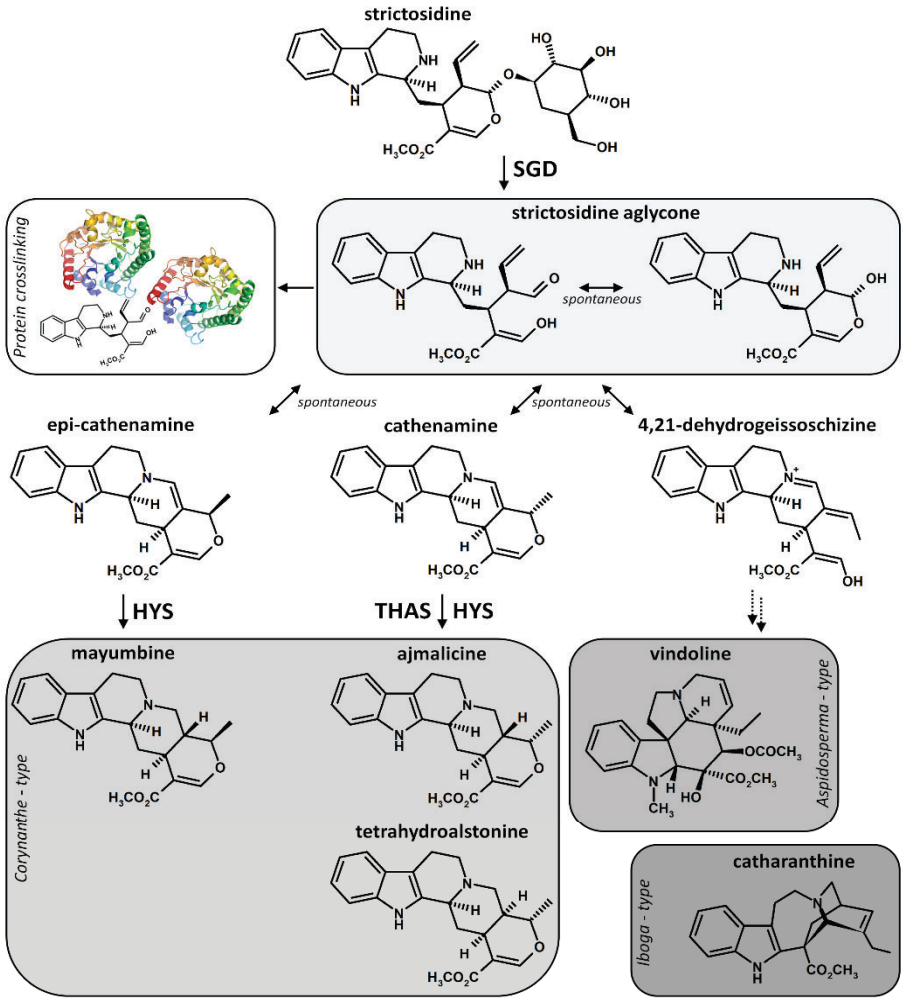


Figure 13: Deglycosylation of strictosidine and associated reactions in *C. roseus*. The deglycosylation of strictosidine by strictosidine β-D-glucosidase (SGD) generates a reactive aglycone that leads to protein crosslinking during the defense process. This may also trigger a spontaneous rearrangement initiating the synthesis of downstream MIAs. The different types of *C. roseus* alkaloids are also presented. Enzymatic and spontaneous reactions are indicated by enzyme name or blank in front of arrows. Double dashed arrows represent multiple biosynthetic reactions. HYS: heteroyohimbine synthase; THAS: tetrahydroalstonine synthase. Figure taken from Carqueijeiro et al. (2021).

Besides the use of highly reactive weapons as the strictosidine aglycon, a good defense mechanism needs to be complete, and that is the reason that leads *C. roseus* to change its complete MIA profile during an attack. That is probably why an accumulation of MIAs including vindoline is observed around the damaged zone, to prevent attacks from herbivores. For the final route alkaloids, in general, it is observed that their metabolism increases significantly during an infection (Vázquez-Flota et al., 2004). In this way, we can observe that different mechanisms of defense rule the alkaloids distribution in *C. roseus*, which can be related to cell compartmentation or metabolism regulation in a case of a biotic stress.

5. Metabolic engineering

5.1. Metabolic engineering applications

Metabolic engineering is the set of rational changes in the genetic material of an organism, altering its metabolic profile or improving its biosynthetic capabilities and it has been increasingly applied in the production of specialized metabolites by heterologous organisms. In the early 2000's, the transfer of genes encoding functional enzymes from plants into genetically mastered hosts such as the bacteria *Escherichia coli* and the yeast *Saccharomyces cerevisiae* started a trend in biotechnology. The modest, however groundbreaking production of natural products via these microorganisms inspired researchers to work on the optimization of this production and today, we have a history of success that continues to bring new perspectives to the field. After the first plant genome sequencing in 2000 (*Arabidopsis thaliana*) (The Arabidopsis Genome Initiative, 2000), we saw in 2003 the production of hydrocortisone from a simple carbon source using the yeast *Saccharomyces cerevisiae* (Szczebara et al., 2003). This amazing achievement was followed by the production of the anticancer Taxol precursors in bacteria (Ajikumar et al., 2010), the great antimalarial artemisinin in yeast (Paddon et al., 2013), *C. roseus* alkaloids as vindoline (Qu et al., 2015) and even cannabinoids were produced recently in yeast (Luo et al., 2019).

The rewarding story of metabolic engineering of natural products still drives the advances in this field, especially with the recent improvement of the molecular toolkit of CRISPR/Cas9 genome editing. In this way, a successful procedure for the exploitation of this technology relies on benchmark pipeline illustrated on the figure 14 (Courdavault et al., 2021).

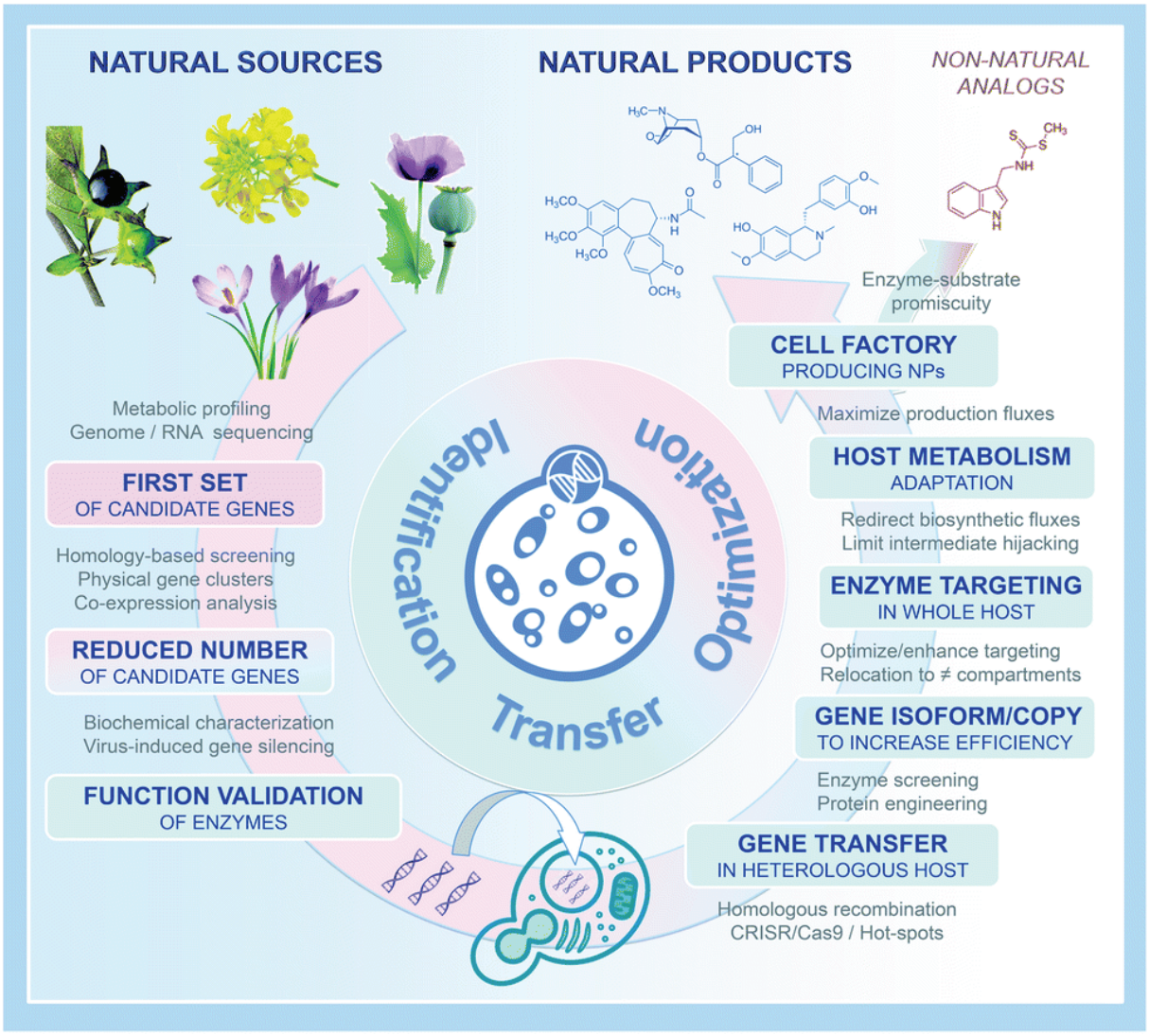


Figure 14: A standard procedure for heterologous organism production of natural products. Figure taken from Courdavault et al. (2021).

5.2. The elucidation of plants natural product biosynthetic pathways

The elucidation of plant biosynthetic pathways requires the use of several techniques that will rely on transcriptomics analysis to select gene candidates that will be further validated. The first step is to have access to the transcripts of the plant of interest. For *C. roseus*, large projects such as Phytometasyn (http://www.phytometasyn.ca) (Xiao et al., 2013) and Medicinal Plant Genomics Consortium (http://medicinalplantgenomics.msu.edu/) (Góngora-Castillo et al., 2012) helped with the fast discovery of MIA biosynthetic genes. Indeed, a project called "epidermome" was essential for the MIA pathway closure in *C. roseus*, by the assembly of transcripts extracted from the leaf epidermis, it opened the access to the hotspot of MIAs on

this plant (Murata et al., 2008). These databases are important to obtain the qualitative raw sequences and the quantitative abundance estimation data, which leads to the clustering and possible comparisons with metabolic data.

Taking a look at the pipeline (figure 15) described by (Dugé de Bernonville et al., 2015), a solid strategy is to use both orthology comparison with other species producing MIAs and clustering of gene expression. That begins with a prior identification of conserved genes between the selected MIA producing species that share a common precursor as the secologanin, for example. This process is followed by a second orthology in which the set of conserved genes is compared with non secologanin producing species, resulting in an even more selected pool of genes.

Another strategy to be put in place is to perform experiments in which the plant of interest will be submitted to different conditions, generating transcriptome data from a diverse of metabolic states of the plant, constructing a reference transcriptome that estimates the abundance of each transcript in a selected condition, clustering the genes to their expression level. The more experimental conditions are performed, the more efficiently the groups will co-regulate. This kind of database is useful not only for elucidating a pathway of interest but also to create a toolbox for metabolic engineering of MIA biosynthesis (Dugé de Bernonville et al., 2015).

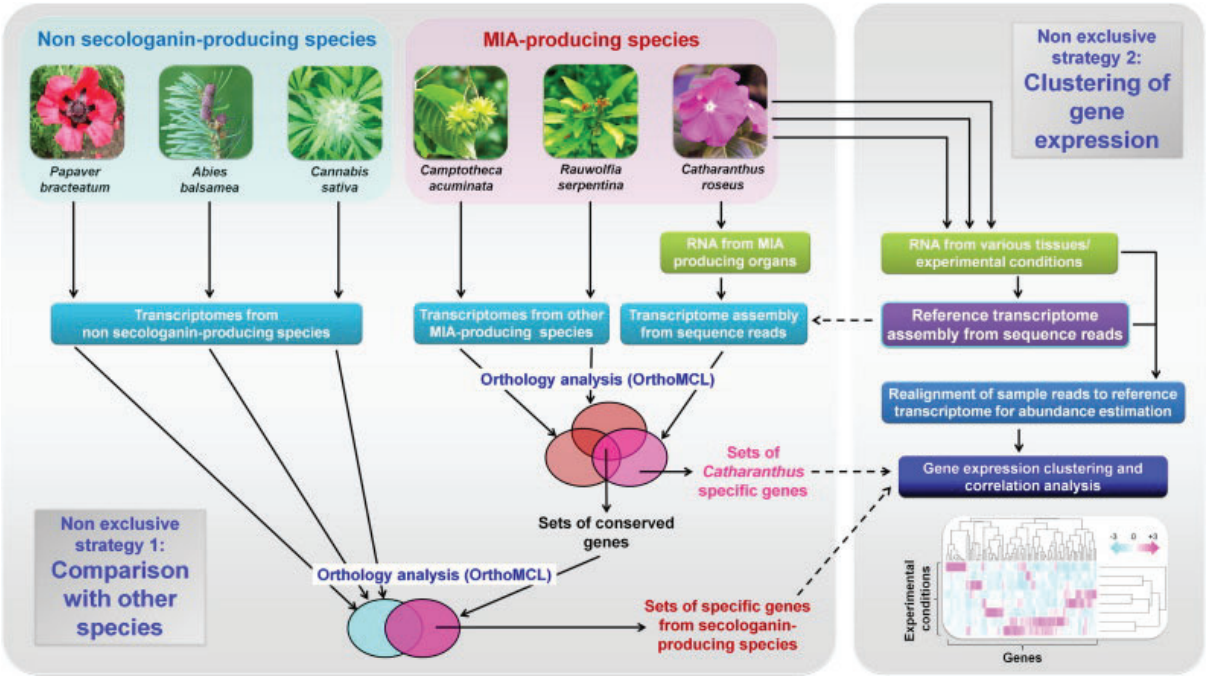


Figure 15: The use of high-throughput transcriptomic data for discovery of MIA biosynthetic genes. Two strategies may be used in order to elucidate MIA biosynthetic genes: 1) orthology comparison with other species (left panel) in which transcriptomic data from candidate species with specific alkaloid metabolisms (e.g. secologanin producing (purple circle) and non secologanin producing (blue circle) species are compared using the OrthoMCL algorithm (Li et al., 2003). 2) Clustering of gene expression (right panel) which helps the identification of conserved genes encoding enzymes involved in the biosynthesis of common precursors in MIA-producing. Analysis from diverse experimental conditions to construct a reference transcriptome which will be used to estimate the abundance of each transcript in each condition in order to cluster genes according to their expression levels. Figure taken from Dugé de Bernonville et al. (2015).

5.3. Improving the mechanisms of gene transfer into the heterologous hosts

A variety of techniques are available nowadays to reach DNA insertion into heterologous organism, including electroporation, conjugation, protoplast transformation and natural competence, for example. Once DNA is inside the cell, it needs to be maintained during cell division. This can be achieved through autonomous replication or through chromosomal integration (Riley & Guss, 2021). Plasmid-based autonomous replication is an effective approach to obtain a high transformation rate in an accessible and fast way. However, for metabolic engineering, this technique can bring problems such as variability between strains,

a high metabolic burden and the fatidic need for continuous antibiotic selection, which is not optimal for an industrial environment.

Chromosomal integration tools such as the homologous recombination approach innovated the field by allowing the introduction of modifications to an organism's chromosome by creating a scar-less mutation, which triggers the DNA repair machinery of the cell and the introduction of the gene of interest, or the Cre/lox-mediated integration (Hochrein et al., 2018). However, these approaches have been largely replaced by the CRISPR/Cas9 system, which uses clustered regularly interspaced palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins (Jinek et al., 2012; Riley & Guss, 2021).

CRISPR-Cas system introduces scarless point mutation, insertions or deletions in the genome. This is possible due to the presence of spacers separated by short repeats in a native CRISPR-Cas system. Normally, these spacers will be filled by invading DNA to enable further attack. The Cas nuclease cleaves the DNA at this sequence guided by the gRNA (guide RNA). In the case of Cas9, the array is transcribed and processed into single RNA (sgRNA), and is the most commonly used nuclease for heterologous system editing. On this case, the sgRNA is transcribed as a single RNA and the double-stranded break created by Cas9 will be repaired using the template containing the desired modification, resulting in an edited genome (Deltcheva et al., 2011; Guirimand et al., 2021; Riley & Guss, 2021).

5.4. Optimization of the biosynthetic pathways up to an industrial level

In order to increase natural products production, an implementation of bioreactors can be done, followed by media and culture optimization such as fed batch, high density cultures, change sugar source, pH and aeration. These modifications can improve the yield of the final product, however molecular modifications of the heterologous host may be required. The most straightforward molecular modifications that have an immediate impact on the productivity are the identification of the production bottleneck that means, to check in which reaction step a poor conversion is achieved. When the non-effective protein is identified, an expression of this enzyme orthologue or isoforms can be done, together with an increase of the copy number of genes encoding the limiting step. Further modifications can be considered, such as co-factor enhancement and protein fusion in order to create a metabolic channeling (Courdavault et al., 2021).

Alternatively, the reconstitution of the pathway can be shortened in order to reach the desired yield. Some long pathways such as the MIA pathway of *C. roseus* shows low yield when build for *de novo* production, which means, from a sugar source, requiring the primary metabolism of the heterologous host to be part of the engineered pathway, such as the strictosidine production (Brown et al., 2015). In this way, a precursor-directed production can be a solution. As previously discussed, the final desired product such as alkaloids, are found in very low concentration in plants (around 2 ppm), while alkaloids precursors such as tabersonine, can be found in high concentrations and can be easily extracted to be fed to a metabolic engineered yeast containing the subsequent genes of the metabolic route. Thus, around 32 genes insertions and deletions will be replaced by only around 7 genes insertions, facilitating the optimization of the process and the identification of possible bottlenecks (Qu et al., 2015).

5.5. Expand the library of molecules by the creation of non-natural functional products

Considering the incredible advances that have already been made in natural product production in heterologous hosts, the idea of using these mechanisms to produce compounds that are completely new to nature emerged. The diversifications by increasing chemical decorations are limitless using the machinery of a heterologous host. A recent example is the halogenation of natural products, which increases the properties of molecules such as alkaloids for possible further derivatization (Courdavault et al., 2021).

An example of this approach using *C. roseus* enzymes was performed by using the enzymes T16H2, 16OMT, T3O, tabersonine 19-hydroxylase (T19H) and tabersonine 19-*O*-acetyltransferase (TAT) (figure 16) (Carqueijeiro et al., 2018). The first two enzymes (T16H2 and 16OMT) are mainly found in leaves of *C. roseus*, whereas the last two (T19H and TAT) are exclusively root proteins. On the first top reaction of the figure 16, the *in planta* reaction found in leaves from tabersonine produces the natural 16-methoxytabersonine epoxide. The second reaction shows the formation of tabersonine epoxide catalysed by T3O in leaves of *C. roseus*, and the formation of 19-hydroxytabersonine catalysed by T19H, a *C. roseus* alkaloid found only in roots. Even when expressed together, both enzymes (T3O and T19H) did not perform sequential decorations. Conversely, we observe that the sequential reaction using the roots proteins (T19H and TAT) can use the leave's product 16-methoxytabersonine as substrate

to produce 16-methoxy, 19-acetyltabersonine that has been never observed in nature. This result demonstrated that a simple enzyme combination from the same plant can orientate the synthetic flux generating non-natural products which could have therapeutic effects.

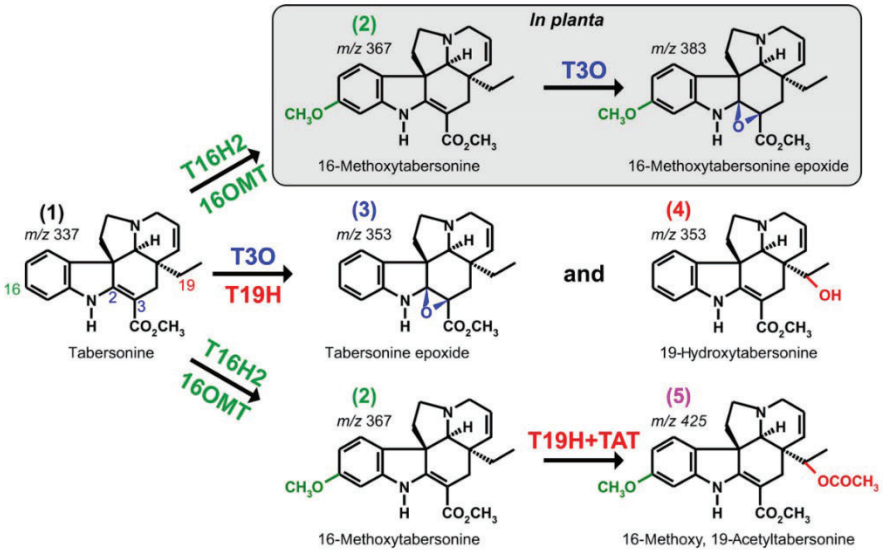


Figure 16: Expansion of the library of molecules by associating different *C. roseus* enzymes. MIA biosynthetic enzymes from roots (T19H, TAT) and leaves (T16H2, 16OMT, T3O) and formation of MIAs. Figure taken from Carqueijeiro et al. (2018).

II. Thesis organisation and objectives

The monoterpene indole alkaloids (MIA) are powerful molecules with remarkable therapeutic activities, such as the vincristine and vinblastine produced by the Apocynaceae plant, *Catharanthus roseus*. Despite the increased economic value of these molecules, the plant remains the main source of purification for medicine purposes, which results in an expensive and laborious task due to the very small concentrations in which they are found naturally. The advancement of biotechnology and genetic tools have shown that alternative productions of similar molecules are already a reality in heterologous organisms. However, the complex architecture of *C. roseus* MIA biosynthetic pathway has brought some challenges that need to be overcome in order to reach the required industrial yield. Fortunately, the MIA pathway have been completely elucidated in *C. roseus*, which allows us to use this massive expansion of transcriptomic resources to search for tools in *C. roseus* or in other species of Apocynaceae family to optimize the MIA production in heterologous organisms. This thesis will show a complete approach to study and produce *C. roseus* MIAs, from gene elucidation to yeast metabolic engineering and optimization.

Part 1 presents the methods for identification and validation of new enzymatic steps of MIA biosynthetic pathways. The article 1 presents the virus-induced gene silencing technique, a tool that allows the study of gene function in *C. roseus* and in *Rauwolfia serpentina*, showing the transposable character of this method in the Apocynaceae family. The article 2 introduces an agrobacteria-mediated overexpression, a complementary method for the functional characterisation of genes in *C. roseus* that allows the protein production *in planta* in order to assess its activity. The article 1 was published on the "Methods in Molecular Biology", while the article 2 was accepted for publication in the same journal.

The part 2 firstly describes the characterisation of a second isoform of the tabersonine-16-*O*-methyltransferase (16OMT) in *C. roseus*. This enzyme is involved in the vindoline production, being of great interest for vinblastine production. The results are presented in the submitted article form (article 3), which is under review by the "Plant Science" journal. The second part presents article 4, the gene identification, by transcriptomic studies, and the characterization of a 16OMT-like enzyme (Vm16OMT) in another Apocyaneae family plant, *Vinca minor*. This enzyme is a 16OMT orthologue and accepts tabersonine derivatives as substrates. The article 4 was published at the "Biomolecules" journal.

Finally, part 3 will present the application of the previous elucidate enzymes of the thesis in the form of two articles describing the optimization of the vindoline pathway in the

yeast *Saccharomyces cerevisiae.* The article 5 introduces molecular tools to reduce the by-product production (vindorosine) and the vindoline intermediates accumulation both using episomal-expression plasmids and stable yeast constructs. The article 6 reports the biggest conversion of tabersonine to vindoline in a bioreactor scale by using a yeast strain constructed *via* CRISPR-cas9 molecular tool. The article 5 was published in the "Molecules" journal, while article 6 was published in the "Microbial biotechnology" journal.

Objectifs et organisation de la thèse en français

Les alcaloïdes indoliques monoterpéniques (AIM) sont des molécules aux activités thérapeutiques remarquables, comme la vincristine et la vinblastine produites par *Catharanthus roseus*, une plante de la famille des Apocinacées. Malgré la valeur économique accrue de ces molécules, la plante reste la principale source de ces molécules utilisées à des fins médicales. Les très faibles concentrations de ces molécules dans la plante sont à l'origine de leur extraction laborieuse et de leurs coûts élevés. Les progrès des biotechnologies et des outils génétiques ont montré que les productions alternatives de molécules similaires par des organismes hétérologues sont déjà une réalité. Cependant, l'architecture complexe de la voie de biosynthèse des AIM de *C. roseus* a posé certains défis qui doivent être surmontés afin d'atteindre une production industrielle. Récemment, la voie de biosynthèse des AIM de *C. roseus* a été complètement élucidée, ce qui nous permet d'utiliser l'expansion massive des ressources transcriptomiques pour rechercher des outils génétiques chez *C. roseus* ou chez d'autres espèces de la famille des Apocynaceae pour optimiser la production d'AIM dans des organismes hétérologues. Cette thèse présente une approche globale pour produire des AIM de *C. roseus*, de l'élucidation des gènes à l'ingénierie métabolique et son optimisation dans la levure.

La partie 1 présente les méthodes d'identification et de validation de nouvelles étapes enzymatiques des voies de biosynthèse des AIM. L'article 1 présente la technique de silençage génique induit par un virus (VIGS: virus-induced gene silencing) qui permet d'étudier la fonction des gènes de *C. roseus* mais aussi de *Rauwolfia serpentina*, montrant le caractère transposable de cette méthode aux plantes de la famille des Apocynaceae. L'article 2 introduit une surexpression médiée par les agrobactéries, une méthode complémentaire à la précédente pour la caractérisation fonctionnelle des gènes de *C. roseus* qui permet la production de la protéine *in planta* afin d'évaluer son activité. L'article 1 a été publié sur « Methods in Molecular Biology », tandis que l'article 2 a été accepté pour publication dans la même revue.

La partie 2 décrit en premier lieu la caractérisation d'une seconde isoforme de la tabersonine-16-O-méthyltransférase (16OMT) de *C. roseus*. Cette enzyme est impliquée dans la production de vindoline, une molécule d'un grand intérêt pour la production de vinblastine. Les résultats sont présentés sous la forme d'un article soumis (article 3), qui est en cours d'examen par la revue « Plant Science ». La deuxième partie présente l'article 4, qui traite de

l'identification de gènes par des études transcriptomiques et de la caractérisation d'une enzyme de type 16OMT (Vm16OMT) chez une autre espèce de la famille des Apocyaneae, à savoir *Vinca minor*. Cette enzyme est un orthologue 16OMT qui accepte des dérivés de la tabersonine comme substrats. L'article 4 a été publié dans la revue « Biomolecules ».

Enfin, la partie 3 présente une application des enzymes élucidées précédemment au cours de la thèse sous la forme de deux articles décrivant l'optimisation de la voie de la vindoline chez la levure *Saccharomyces cerevisiae*. L'article 5 présente des outils moléculaires pour réduire la production de sous-produits (vindorosine) et l'accumulation d'intermédiaires de la vindoline à la fois en utilisant des plasmides d'expression et des constructions stables dans la levure. L'article 6 présente la plus grande bioconversion de tabersonine en vindoline obtenue à ce jour dans un bioréacteur en utilisant une souche de levure transformée avec l'outil moléculaire CRISPR-cas9. L'article 5 a été publié dans la revue « Molecules », tandis que l'article 6 a été publié dans la revue « Microbial biotechnology ».

III. Results

Part 1 : Methods used for the identification and validation of biosynthetic genes

This first part of the results will present two articles describing the methods used along the thesis for the identification and validation of candidate genes *in planta.* Both articles are of extreme importance for the domain of biotechnology due to its application for enzymatic steps of the metabolic pathways that were not yet elucidated. This knowledge is essential when re-constructing a plant metabolic pathway in a heterologous organism, even more so to reach its complete optimization. Lastly, these techniques are transferable to the study of other metabolic pathways in other plant species.

The virus-induced gene silencing (VIGS) and the agroinfiltration-mediated transient overexpression (OE) techniques will be presented on this part. Despite the differences between the two methods, they are complementary and should be performed in combination to gauge the function of a specific gene. In short, the VIGS technique is a RNA based technique that takes advantage of endogenous plant defence mechanisms in order to downregulate a specific plant gene and assess its function through its metabolic response. On the other hand, the OE technique makes use of the *Agrobacterium tumefaciens*, a bacteria which possess the ability to transfer a target DNA sequence to the plant cell nucleus, resulting in the expression of the recombinant proteins by the plant. This ability can be explored in order to overexpress a plant protein and equally inquire about its function by evaluating the metabolic response.

In terms of pros and cons, the VIGS technique does not rely on the generation of stable genetic transformants and a minimum input of time and effort is required so many lines/cultivars can be compared in a short time. Differently, OE requires the construction of *Agrobacterium tumefaciens* strains harbouring the vectors of interest and also call for prior culture preparation and virulence induction. A further advantage of the OE is the possibility to deliver several transgenes into the same cells (Kapila et al., 1997) so that multimeric proteins, such as antibodies, can be expressed and assembled (Vaquero et al., 1999).

Another aspect to consider is that VIGS is based on a sequence-specific mechanism, which means that both single copy genes and multiple members of a multigene family can be silenced in the same time (Pflieger et al., 2013). This implies that cross-silencing can be observed depending of the sequences chosen for the silencing. In the VIGS approach, gene fragments should not exceed 300 bp, which makes the choice of the best region for silencing

harder when working with similar genes, such as gene encoding for protein isoforms. The OE approach could be beneficial when working with similar sequences due to the possibility of using >100 kbp sequences, increasing the specificity of the validation of the target gene (Whitham et al., 2016). The VIGS and the OE approaches represents two valuable and complementary approaches to study the Apocynacea plants metabolisms in regards to its own particularities.

Article 1: A biolistic-mediated virus-induced gene silencing in Apocynaceae to map biosynthetic pathways of alkaloids

This publication is a method guide put in place to identify candidate genes by considering the metabolic profile of the plants when a target gene is silenced by Virus-induced gene silencing (VIGS). Particularly, this approach constitutes a simple and effective alternative to the elucidation of plant gene function by taking advantage of post-transcriptional gene silencing (PTGS) mechanism occurring naturally in plants upon viral infections (Sung et al., 2014). Through a specific downregulation of a target plant gene, the VIGS technology allows the knockdown of a certain phenotype that will be used to identify the gene's function.

In summary, a basic RNA silencing pathway follows 4 steps: (1) dsRNA (double strand RNA) or hpRNA (hairpin structure RNA) are processed by the Dicer or Dicer-like proteins (DCL) into 20_24 nucleotide (nt) small RNA (sRNA) duplex with 2-nt 3' overhangs at both ends. (2) One strand from the sRNA duplex is integrated into the Argonaute (AGO) protein, resulting in the formation of an RNA-induced silencing complex (RISC). (3) The RISC complex will be guided by the sRNA to its complementary region of single stranded RNA. (4) Once the sRNA finds its homologous mRNA and the binding occurs, the AGO protein will chop the mRNA, inhibiting its translation (Guo et al., 2016).

The VIGS technique takes advantage of this specific mechanism by using a viral genomic RNA of a replication-competent virus (usually *Tobacoo rattle virus*) in order to trigger the sequence-specific degradation of RNA, by the formation of double-stranded RNA (Hammond, 2001). The elucidation plant gene's function will be achieved by constructing a recombinant virus carrying the partial sequence of the gene of interest (GOI) (preferably more than 23 base pairs) and using it to infect the plant. This infection can be done in several ways such as *Agrobacterium* infiltration or biolistic delivery. The viral gene transcripts (mRNAs) will be recognized and degraded along with the target plant mRNAs by the endogenous system previously described (PTGS), resulting in a downregulation of the target gene expression (Yan et al., 2018).

In the article presented, the VIGS method is explained using the gene geissoschizine oxidase as an example. This gene participates in central point reactions of the MIA pathway of *Catharanthus roseus*, being responsible for the oxidation of geissoschizine. In this work, I was responsible for the conduction of VIGS experiments for both *C. roseus* and *R. serpentina*. In

addition, I was also responsible for the writing of the article with the assistance of a post doc researcher which developed the method.

RESUME EN FRANÇAIS DE L'ARTICLE 1

Les alcaloïdes indoliques monoterpéniques (AIM) sont des métabolites spécialisés synthétisés dans de nombreuses plantes de la famille des Apocynaceae comprenant *Catharanthus roseus* et *Rauvolfia* sp. Les AIM font partie de l'arsenal chimique des plantes pour faire face aux attaques des animaux et des herbivores. Leurs activités biologiques leur confèrent également des propriétés pharmaceutiques qui sont exploitées en pharmacopée humaine. Le développement robuste et simple d'outils pour élucider chaque étape des voies de biosynthèse des AIM constitue ainsi un préalable à la compréhension des mécanismes de défense des Apocynaceae et à l'exploitation de la cytotoxicité des AIM en les produisant par ingénierie métabolique. Alors que les protocoles de silençage génique induit par un virus (VIGS: Virus-Induced Gene Silencing) et basés sur des transformations via Agrobacterium ont émergé, la récalcitrance des Apocynaceae à ce type de transformation nous a incité à développer une procédure universelle d'inoculation basée sur la méthode VIGS. Une telle procédure repose sur la délivrance de plasmides transformants via un dispositif de bombardement biolistique qui offre la possibilité de surmonter la spécificité de l'hôte de n'importe quelle espèce végétale. En utilisant comme exemple l'invalidation du gène codant la geissoschizine oxydase, nous avons décrit les principales étapes de cette technique biolistique chez *C. roseus* et *R. tetraphylla*.



Chapter 8

A Biolistic-Mediated Virus-Induced Gene Silencing in Apocynaceae to Map Biosynthetic Pathways of Alkaloids

Pamela Lemos Cruz, María Isabel Restrepo, Thomas Dugé de Bernonville, Audrey Oudin, Thibaut Munsch, Arnaud Lanoue, Sébastien Besseau, Lucia Atehortùa, Nathalie Giglioli-Guivarc'h, Nicolas Papon, Marc Clastre, Inês Carqueijeiro, and Vincent Courdavault

Abstract

Monoterpene indole alkaloids (MIAs) are specialized metabolites synthesized in many plants of the Apocynaceae family including *Catharanthus roseus* and *Rauvolfia* sp. MIAs are part of the chemical arsenal that plants evolved to face pet and herbivore attacks, and their high biological activities also confer pharmaceutical properties exploited in human pharmacopeia. Developing robust and straightforward tools to elucidate each step of MIA biosynthetic pathways thus constitutes a prerequisite to the understanding of Apocynaceae defense mechanisms and to the exploitation of MIA cytotoxicity through their production by metabolic engineering. While protocols of virus-induced gene silencing (VIGS) based on *Agrobacterium*-based transformation have emerged, the recalcitrance of Apocynaceae to this type of transformation prompted us to develop an universal procedure of VIGS vector inoculation. Such procedure relies on the delivery of the transforming plasmids through a particle bombardment performed using a biolistic device and offers the possibility to overcome host specificity to silence genes in any plant species. Using silencing of geissoschizine oxidase as an example, we described the main steps of this biolistic mediated VIGS in *C. roseus* and *R. tetraphylla*.

Key words VIGS, Silencing, Apocynaceae, Alkaloids, Gene validation

1 Introduction

From the analgesic morphine produced by *Opium poppy* to the antiasthma ephedrine extracted from *Ephedra* spp., the bioactive compounds produced by plants are countless. The capacity of plants to architecture and decorate molecules goes beyond our own imagination such as in Apocynaceae producing the valuable monoterpene indole alkaloids (MIAs). Remarkable plants of this family first include the Madagascar periwinkle (*Catharanthus roseus*), source of vinblastine and vincristine, two MIAs that

Vincent Courdavault and Sébastien Besseau (eds.), *Virus-Induced Gene Silencing in Plants: Methods and Protocols*, Methods in Molecular Biology, vol. 2172, https://doi.org/10.1007/978-1-0716-0751-0_8,
© Springer Science+Business Media, LLC, part of Springer Nature 2020

constitute the first natural drugs used in chemotherapy, with special emphases on the most severe types of cancer such as Hodgkin's lymphoma and lung and brain cancer. It also encompasses *Rauwolfia* sp. such as *R. serpentina* and *R. tetraphylla*, known as the devil pepper, which both synthesize the antiarrhythmic ajamaline accumulated in stem barks and roots but frequently substituted by the semisynthetic propyl derivative—prajmaline—due to its low availability [1, 2].

Over the last five decades, unraveling the metabolic pathways of MIAs in those plants constituted an outstanding scientific topic that challenged numerous researchers of the specialized metabolism field. Besides deciphering the role of specialized metabolisms in plant/environment interactions, it allowed understanding how plants recruited enzymes from primary metabolisms to synthesize secondary metabolites. Nowadays, such elucidations find a renewed interest with the last developments in synthetic biology allowing us to consider other "biological factories" for the production of these valuables molecules, including genetically modified yeast or bacteria [3–5]. Developing new approaches to produce compounds of interest constitutes a concrete alternative to the classical supply relying on extraction from plants notably for MIA production. At midterms, it may provide an answer to the potential supply problems that may come from natural resources endangering or climate changes, as illustrated with the anti-malaria artemisinin, previously extracted from *Artemisia annua*, which can now be obtained through the synthesis of its precursor artemisinic acid by both types of engineered microbes [3]. However, conferring to other organisms the capacity to produce compounds of interest on demand relies on the transfer of the ad hoc producing machineries and thus on the identification of each gene of the corresponding biosynthetic pathways.

Undeniably, development of OMIC's approaches and especially RNA-seq constitutes a turning point in biosynthetic gene identification by providing access to the sequences of all the genes simultaneously expressed in the different cells/organs of plants. Gathering gene sequences co-expressed in organs accumulating valuable molecules gives a potential access to candidate genes involved in their biosynthesis and in turn rises the demand on characterization tools to understand the role of these genes. Although mutant plants generated by chemical mutagenesis, T-DNA tagging [6–9], or RNA interference [10, 11] gave potent results, the laborious and time-consuming task of mutant population characterization [12] promotes the development of new strategies. In this context, transiently silencing genes through virus-induced gene silencing (VIGS) rises a major breakthrough, conferring the capacity of evaluating the consequences of gene downregulation in short time and thus quickly elucidating the function of numerous candidate genes [13].

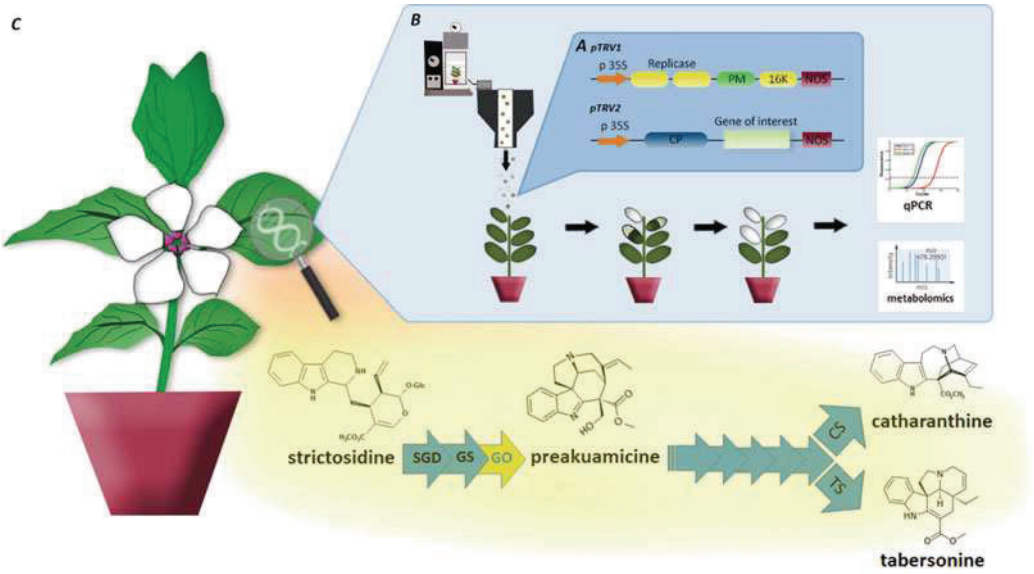


Fig. 1 Flowchart of TRV-mediated VIGS of the gene geissoschizine oxidase in *C. roseus*. From the viral vectors (**a**), the biolistic delivery (**b**) to the metabolic pathway (**c**)

From a technical point of view, VIGS deeds the capacity of plants' defense mechanisms to degrade double-stranded RNAs formed during invading virus replication [14]. It thus belongs to posttranscriptional gene-silencing approaches [14, 15]. Taking advantages of this process, VIGS involves the cloning of the viral component(s) in (binary) vectors to allow delivery in plants, under the control of a strong promoter such as CaMV35S for initial constitutive expression (Fig. 1a). A small fragment of the target plant gene (200–500 bp for instance) is inserted in the virus genome causing the degradation of the corresponding plant mRNA through virus RNA targeting by plant defense mechanisms [16–18]. Due to virus host specificities, different VIGS vectors based on distinct virus types have been developed over years expending the spectrum of plants potentially subjected to VIGS. Currently, VIGS vectors based on *Tobacco rattle virus* (TRV) are by far the most widely used. This is notably due to the high capacity of tissue infection including meristems with weak side effects [16, 19] and to the large array of TRV host plants ranging from *Arabidopsis* [20], *Solanaceae* sp. [17, 21], or more recalcitrant species such as plants from the Apocynaceae family [22, 23].

Delivery of the infectious material (from plasmids, RNA, to reformed viral particles) represents a critical step in the initiation of the VIGS procedure. Depending on virus and assayed plants, numerous techniques have been developed. For example, viral vectors can thus be transfected in plant cells mechanically [24], inoculated by *Agrobacterium* through vacuum/syringe leaf

infiltration or by making holes with toothpick [13, 25] or delivered by a biolistic procedure [22, 23, 26]. Albeit more expensive and time-consuming, biolistic-mediated VIGS that involves the bombardment of plants with microparticles coated with viral vectors (Fig. 1b) presents a unique benefit relying on overcoming the host specificity of *Agrobacterium* resulting in the capacity to deliver transforming DNA/RNA in all plants. Besides being highly reproducible, it does not suffer from the pronounced defense reaction that plants may deploy after *Agrobacterium* infection [27]. Therefore, the biolistic-mediated inoculation of viral vectors constitutes a potent approach to perform VIGS in all plant species once efficient viruses identified for the targeted plant.

In the present report, we describe an efficient protocol of biolistic-mediated VIGS initially developed for the medicinal plants of the Apocynaceae family including *C. roseus*, *R. tetraphylla*, or *R. serpentina* and that can be transposed to many plants following minor adaptations. To describe protocol conduct, we illustrated how using this method we were able to uncover the missing enzymes catalyzing the ultimate steps to tabersonine and catharanthine biosynthesis, a story with more than 60 years that reveals the 31 enzymatic reactions needed to complete the anticancer MIA production [28, 29]. Through silencing of geissoschizine oxidase—GO (Fig. 1c)—in *C. roseus* leaves and using phytoene desaturase (PDS) as a reporter gene to visualize VIGS efficiency by leaf photobleaching, we highlighted the crucial steps of the whole silencing process from viral vector delivery up to analyses of gene silencing consequences.

2 Materials

2.1 Plant Materials

1. Seeds of *Catharanthus roseus* (Little Bright Eye) and *Rauwolfia tetraphylla*.
2. Greenhouse under controlled 16-h light/8-h dark cycle with white fluorescent light (intensity of 70 μmol/m^2/s).

2.2 Cloning in pTRV2

1. Liquid nitrogen.
2. Reverse transcriptase kit.
3. Micropipette set.
4. Sterile filter pipette tips.
5. PCR tubes.
6. Sterilized Milli-Q water.
7. 10 mM dNTP Mix.
8. Specific targeted gene oligonucleotides for PCR:

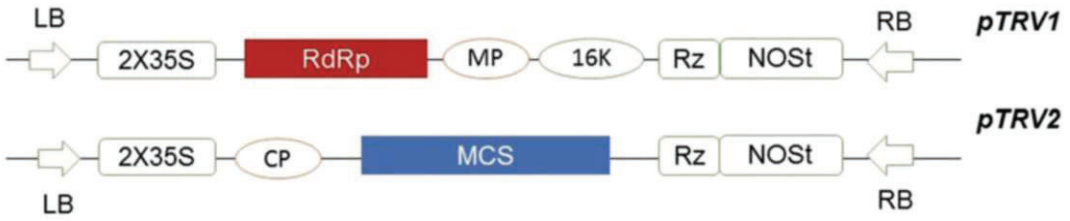


Fig. 2 *Tobacco rattle virus* (TRV)-based VIGS vectors. TRV cDNA clones were placed between duplicated 35S promoter (2×35S) and nopaline synthase terminator (NOSt) in a T-DNA vector. *RdRp* RNA-dependent RNA polymerase, *16K* 16 kDa cysteine-rich protein, *MP* movement protein, *CP* coat protein, *LB and RB* left and right borders of T-DNA, *Rz* self-cleaving ribozyme, *MCS* multiple cloning sites

(a) PDS *C. roseus* forward: 5′-AGGTTTGGGGGGTTTGTG T-3′; reverse: 5′-TACGCCTTGCTTTCTCATCC-3′.

(b) PDS *R. tetraphylla* forward: 5′-AGGTTTGGGGGG TTTGTGT-3′; reverse: 5′-TACGCCTTGCTTTCTCAT CC-3′.

(c) GO *C. roseus* forward: 5′-CTGAGAGGATCCTACAG TATGGCCCGA-3′; reverse: 5′-CTGAGAGGATC-CATCGTTAACAAGATGAGGAACCAAT-3′.

9. High-fidelity DNA polymerase.
10. Thermal cycler.
11. 50× TAE Buffer: 2 M Tris, 1 M glacial acetic acid, 50 mM EDTA, and pH 8.0.
12. 1% agarose gel: 1 g of agarose in 100 mL of 1× TAE buffer and 2 μg/mL ethidium bromide.
13. DNA electrophoresis equipment.
14. 6× Loading buffer: 0.25% (w/w) bromophenol blue, 40% (w/v) sucrose, and 0.25% (w/v) xylene-cyanol.
15. Spectrophotometer.
16. Gel extraction and PCR cleanup kit.
17. RNA purification kit.
18. TRV vectors, namely, pTRV1 and pTRV2-MCS, encoding the two genomic components of TRV (Fig. 2) (Source: Arabidopsis Biological Resource Center—ABRC; http://www.arabidopsis.org).
19. Restriction enzymes and buffers (including *BamH*I).
20. T4 DNA ligase and buffer.
21. *Escherichia coli* strain TOP10 chemically competent.
22. Luria-Bertani Broth: 10 g tryptone, 5 g yeast extract, and 10 g NaCl. Combine reagents, and dissolve in 800 mL deionized water, adjust pH to 7 using 1 M NaOH, and finally adjust volume to 1 L.

23. Luria-Bertani agar: 10 g tryptone, 5 g yeast extract, and 10 g NaCl. Combine reagents, and dissolve into 800 mL deionized water; adjust pH to 7 using 1 M NaOH. Add 20 g of agar and finally adjust volume to 1 L.
24. 50 mg/mL Kanamycin stock solution in sterilized Milli-Q water.
25. Incubator shaker at 37 °C, 200 rpm.
26. Centrifuge.

2.3 Particles Bombardment Procedure

1. Gold particles 1 μm.
2. 0.1 M Spermidine stock solution (*see* **Note 1**).
3. 2.5 M $CaCl_2$ stock solution.
4. Vortex.
5. Absolute ethanol (100%) and ethanol 70%.
6. Sonic bath.
7. Set of micro- and macro-carriers for Bio-Rad PDS1000/He delivery system Bio-Rad (1100-psi rupture disks).
8. Bio-Rad PDS1000/He delivery system Bio-Rad device.
9. Box for plants with acclimatization control.

2.4 qPCR Analysis

1. Quantitative PCR instrument.
2. Microcentrifuge.
3. qPCR SYBR Green Mix.
4. cDNA synthesis kit (reverse transcriptase).
5. RNA purification kit.
6. Oligonucleotides for qPCR:
   (a) qPDS *C. roseus* forward 5′-ATGCCCGTTGTTGATCA-TATTCGAT-3′; reverse 5′-ATCTCCTTTGATTGCTG ACCCATTA-3′.
   (b) qGO *C. roseus* forward: 5′-GCTGAGTTTATGTTGG CTGCTATGTT-3′; reverse: 5′-ATAGTTGGCAAAG ACAGACTAATCGT-3′.
   (c) qRSP9 *C. roseus* forward 5′-TTGAGCCGTATCAG AAATGC-3′; reverse 5′-CCCTCATCAAGCAGACCA TA-3′.
   (d) qPDS *R. tetraphylla* forward 5′-GATCGAGTAACTG ATGAGGTATTCA-3′; reverse 5′-AGCATGGTTCC AAAATGGCCTTTTTA-3′.
   (e) qActin *R. tetraphylla* forward 5′-TGACCTTGAAG-TACCCTATTGAGC-3′; reverse 5′-TGTACGACCAC TGGCATACAGAG-3′.
7. Sterile filter pipette tips.

8. Sterile 1.5 mL screw-top microcentrifuge tubes.
9. PCR tubes or 96 well plates with sealing films.
10. Sterilized Milli-Q water.
11. Microcentrifuge.

2.5 Alkaloid Analysis

1. Mixer mill.
2. Methanol supplemented with 0.1% of formic acid.
3. 90:10 (v/v) of water/acetonitrile (0.1% formic acid).
4. UPLC chromatography system coupled to a SQD mass spectrometer equipped with an electrospray ionization (ESI) source controlled by MassLynx 4.1 software (Waters, Milford, MA) with a C18 column (150 × 2.1 mm, i.d. 1.8 μm) at a stabilized flow rate of 0.4 ml/min at 55 °C.
5. The following multistep gradient was used: acetonitrile (B) versus water (A) system containing 0.1% (v/v) formic acid, at a flow rate of 0.1 mL/min, with 10:90:0.1 to 60:40:0.1 over 18 min.
6. QuanLynx™ software (Waters, UK).

3 Methods

3.1 Plant Germination and Growth Condition Pre-transformation

1. Seeds of *C. roseus* and *R. tetraphylla* are germinated at 28 °C in loam substrate, in greenhouse.
2. Once cotyledons are fully developed, plantlets are transferred to individual pots and cultivated in the same conditions until transformation.
3. For the VIGS procedure, the exact developmental stage of plantlets depends on plant species. For *C. roseus* and *R. tetraphylla*, the plants must have the first pair of leaves full developed and the second pair just emerged as previously accurated for both species [22, 23] (*see* Fig. 3; **Note 2**).

3.2 cDNA Synthesis of Target Genes

1. The total RNA from *C. roseus* and *R. tetraphylla* was isolated from leaf tissue using commercial kit following the instructions of the manufacturer.
2. Quantify total RNA with a spectrophotometer at 260/280 nm (*see* **Note 3**).
3. Analyze the integrity of the RNA by migration on 1% agarose gel.
4. The cDNA synthesis was performed with a cDNA synthesis kit from 1 μg total RNA primed with random hexamers (5 μM), following the manufacturer's instructions (*see* **Note 4**).

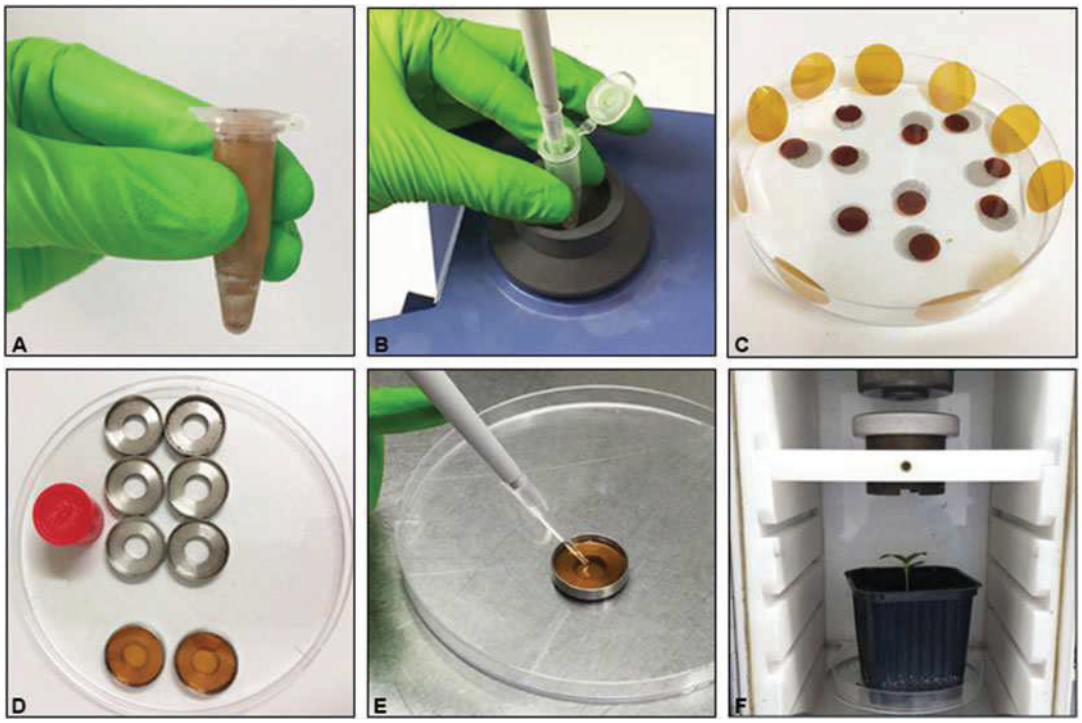


Fig. 3 Pre-bombardment procedure. (**a**) Glycerol-stocked gold particles. (**b**) Coat plasmid constructs onto gold particles. (**c**) Set of macro-carriers, microcarriers, and stopping screens from Bio-Rad used on the transformations. (**d**) Macro-carriers placed on the O-rings. (**e**) Gold-coated particles in 100% ethanol suspension being placed in the macro-carriers. (**f**) *C. roseus* plantlet placed in the Bio-Rad delivery system before transformation (stopping-screen-to-target distance of 9 cm)

5. We considered the sequence (3′-end fragments) of *C. roseus* PDS gene (GenBank accession #JQ655739), the ortholog *R. tetraphylla* PDS gene (similar to rsa_locus_3770_iso_7_-len_2441_ver_2), and GO (CYP71D1V1 GenBank accession #JN61301) as genes that will be silenced (*see* **Note 5**).
6. Amplify DNA fragments using high-fidelity DNA polymerase and the gene-specific oligonucleotides, incorporating the *BamH*I restriction site at both cDNA extremities (*see* **Note 6**).
7. The PCR products must be analyzed by agarose gel electrophoresis to ensure the correct amplification.
8. Purify the PCR product using a PCR cleanup kit.
9. Digest each amplified DNA fragment with *BamH*I.
10. Purify the product of the digestion using a PCR cleanup kit.

3.3 Silencing VIGS Constructs

1. Linearize the pTRV2 vector (around 1 μg) by digestion with the proper restriction enzymes, in this case, *BamH*I.
2. Purify the digestion product using a PCR cleanup kit.

3. Ligate the digested DNA fragments (around 100 ng) into the vector pTRV2 (around 100 ng) to create the final transformation vector (*see* **Note 7**).
4. Transform *E. coli* (strain TOP10) chemically competent cells with the product ligation following a standard heat shock protocol.
5. Select positive clones on Luria-Bertani agar media containing 50 μg/mL kanamycin through standard colony PCR method.
6. Extract the vector using a plasmid extraction kit. Supercoiled plasmids used for particle bombardment should be isolated from *E. coli* desirable yield and quality of the DNA, approx. 1 μg/μL (*see* **Note 8**).
7. Confirm the presence of the insert and its orientation by restriction digestion followed by DNA sequencing.

3.4 Particles Bombardment Procedure

3.4.1 Particle Preparations

1. 30 mg of 1 μm gold particles are weighted in a glass tube followed by dry heat at 180 °C for 8 h (*see* **Note 9**).
2. From this point on, all manipulations should be performed in a flow hood under sterile conditions. Add 1 mL of fresh 70% ethanol, and wash the gold particles for 5 min trough vortex and sonication in an ultrasonication bath. After 5 min, transfer the gold particles to a 1.5 mL sterile microcentrifuge tube (*see* **Note 10**).
3. Pellet the golden particles by centrifugation at 16,000 × *g* for 5 s.
4. Without disturbing the pellet, remove the supernatant.
5. Slowly add 1 mL of sterilized Milli-Q water to wash the pellet. Vortex, and sonicate for 2 min. Pellet the golden particles by centrifugation at 16,000 × *g* for 5 s, and remove the supernatant. Repeat this step two more times. Take side of a P200 pipette to eliminate the remaining water.
6. After washing the golden particles, resuspend in 500 μL of sterilized 50% glycerol (v/v) (*see* **Note 11**).

3.4.2 Coating of Plasmid Constructs onto Golden Particles

1. For each plant to be transformed, prepare a mix of 1 μg of plasmid constructs (pTRV1 and pTRV2—empty vector or harboring a target gene).
2. Add to a sterilized microcentrifuge tube 10 μL of glycerol stocked gold particles (*see* **Note 12**).
3. Mix 5 μL of 0.1 M spermidine with the gold particles while vortexing. Incubate for 3 min at the bench. Keep vortexing for short pulses during the reaction period (Fig. 3a, b).

4. Add the appropriate amount of mixed plasmids to the particles while vortexing. Wait for 3 min, and keep vortexing for short pulses during the reaction period (*see* **Note 13**).
5. Add 6 μL of 2.5 M $CaCl_2$ to the mix while vortexing to optimize the coating, and mix for additional 15 min with a constant speed vortex (1000 rpm).
6. Pellet the gold-coated particles at 16,000 × *g* for 2 s, and remove the supernatant.
7. Wash the pellet with 500 μL of freshly prepared 70% ethanol without disturbing the pellet. Repeat washing using 500 μL of 100% ethanol. Carefully, remove all the remaining liquid near the pellet.
8. Resuspend the gold-coated particles in 10 μL of 100% ethanol with the help of a vortex.
9. Distribute 10 μL of the particles onto the macro-carrier, and let it dry before starting the transformation (Fig. 3c–e).

*3.4.3 Particles Bombardment Procedure (See **Note 14**)*

1. Bombardment of *C. roseus* and *R. tetraphylla* plantlets are performed with the Bio-Rad PDS1000/He delivery system following the manufacturer's recommendations, using 1100-psi rupture disks and under a vacuum pressure of 27.5 in. of Hg.
2. A single parameter of the transformation conditions changes between the two varieties of plants due to their different size, *R. tetraphylla* being around twofold taller than *C. roseus*. Therefore, to maintain a similar distance between top leaves and stopping screen, the stopping-screen-to-target distance was adjusted at 9 and 6 cm, for *C. roseus* and *R. tetraphylla*, respectively.
3. A single bombardment per individual potted plantlet was performed.
4. Transformed plants were kept in greenhouse in a wet atmosphere during 2 days post-bombardment (dpb) followed by progressive reacclimation (Fig. 3f).
5. Plants are cultivated in a greenhouse until appearance of the phenotype.

3.5 Post-transformation Treatments and Analysis

The photobleaching of leaves caused by PDS silencing in the control plants is used to evaluate VIGS efficiency assuming that silencing of the target gene evolves similarly (Fig. 4). For *C. roseus*, leaf photo-bleaching typically begins around 7–10 days after bombarding, while it reaches completion after 21–25 days in neo-formed leaves.

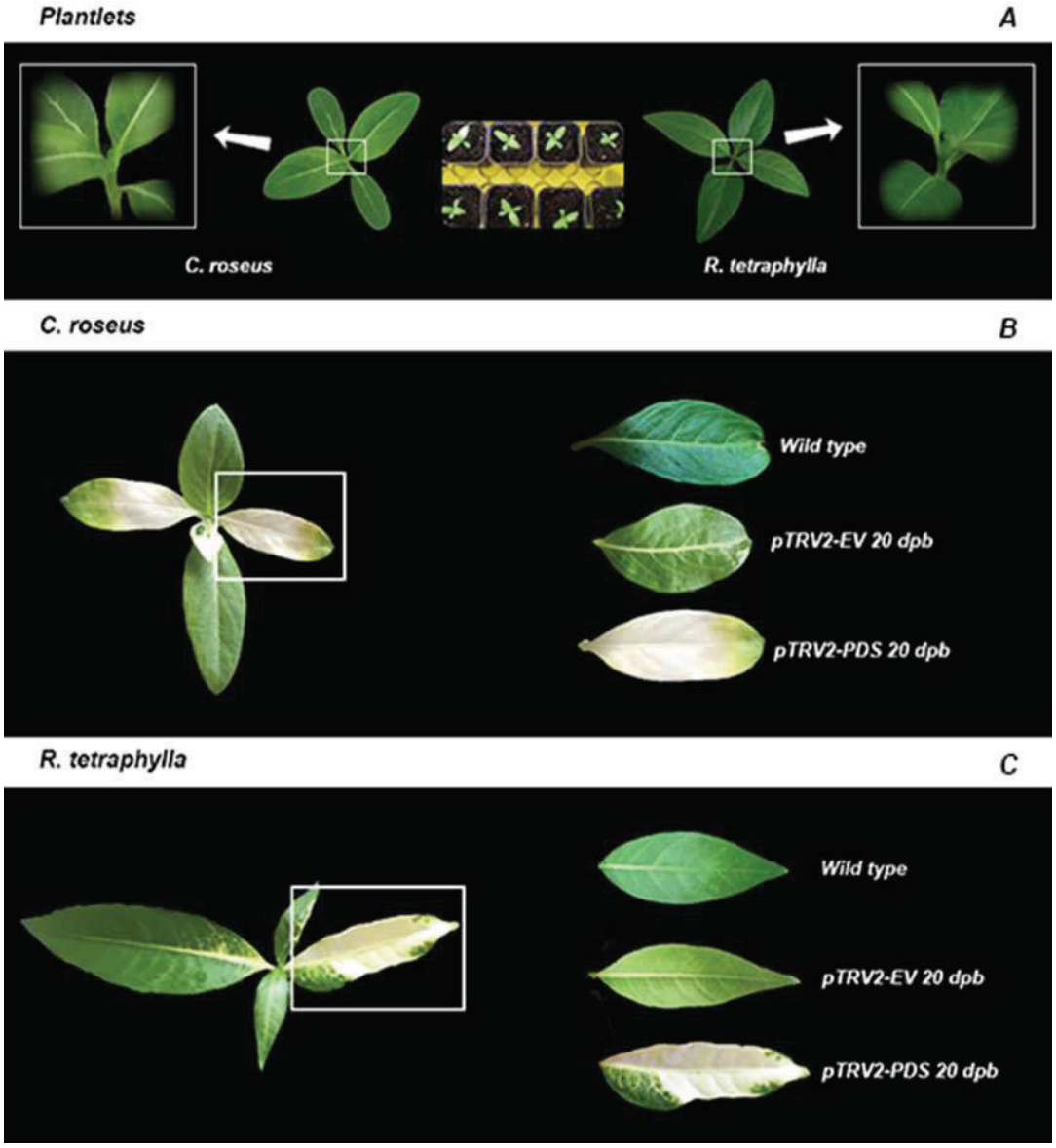


Fig. 4 *C. roseus* and *R. tetraphylla* plantlets transformation. (**a**) *C. roseus* and *R. tetraphylla* plantlets pretransformation presenting one pair of fully expanded leaves. (**b**) *C. roseus* plantlet 20 days after bombardment (dbp) with pTRV2-PDS and the leaves from the wild-type plantlet; plantlet transformed with an empty vector (pTRV2-EV) or with the silencing PDS construct (pTRV2-PDS), all at 20 dpb. (**c**) *R. tetraphylla* plantlet 20 dbp with pTRV2-PDS and the leaves from the wild-type plantlet; plantlet transformed with an empty vector (pTRV2-EV) or with the silencing PDS construct (pTRV2-PDS), all at 20 dpb

3.5.1 Harvesting of Silenced Leaves

1. Collect the first two leaf pairs to emerge following inoculation.
2. In advance, prepare Safe-Lock sterilized microcentrifuge tubes containing three glass beads (*see* **Note 15**).
3. Weight the tubes.

4. Collect the two young leaves, one destined to RNA extraction for testing gene silencing by qPCR and the other one designed for MIA extraction. Flash-freeze the samples (*see* **Note 16**).
5. Store the samples at −80 °C upon use.

3.5.2 Quantitative PCR Analyses

1. Isolate total RNA from the silenced leaves collected.
2. Quantify total RNA with a spectrophotometer at 260/280 nm (*see* **Note 17**).
3. Analyze the integrity of the RNA by migration on 1% agarose gel.
4. Perform cDNA synthesis using 1 μg of total RNA using the reverse transcriptase kit.
5. Before starting the qPCR reactions, place all needed components on ice, vortex briefly each one, and spin down during 2 s to collect the contents at the bottom of the tube.
6. Prepare a master mix according to the protocol of the chosen SYBR green qPCR kit. Always add 10% volume to allow for pipetting error (*see* **Note 18**). As an example, a single reaction is contained in a 15 μL final volume, 6 μL of diluted template cDNA (18×), 0.5 μM forward, and reverse primers designed for the targeted gene and 2× SYBR green Mix.
7. To validate the efficiency of target gene silencing (GO), use in separate reactions primers designed for the targeted gene and primers for housekeeping gene (RPS9 for *C. roseus* or Actin for *R. tetraphylla*) (*see* **Note 19**).
8. Transfer the appropriated volume of master mix to each well on the plate or PCR tube.
9. Cover the PCR plate or the PCR tubes and centrifuge it briefly.
10. Program the qPCR protocol: 95 °C for 7 min (Initial denaturation), 40 cycles at 95 °C for 10 s (denaturation), and 60 °C for 40 s (annealing, extension, and read fluorescence). Hold at 4 °C is optional.
11. The results of the qPCR are normalized in comparison with a housekeeping gene using ΔΔCt method.

3.5.3 Alkaloids Extraction Protocol (See Note 20)

1. After 24 h at −80 °C, lyophilize the samples dedicated to MIA extraction for 24–48 h.
2. Grind the samples with mixer mill during 3 min at the maximal frequency.
3. Weight the tubes and determine the precise dry weight mass of each sample.
4. Add 1 mL of methanol: 0.1% formic acid solution.
5. Vortex during 1 h at 24 °C.

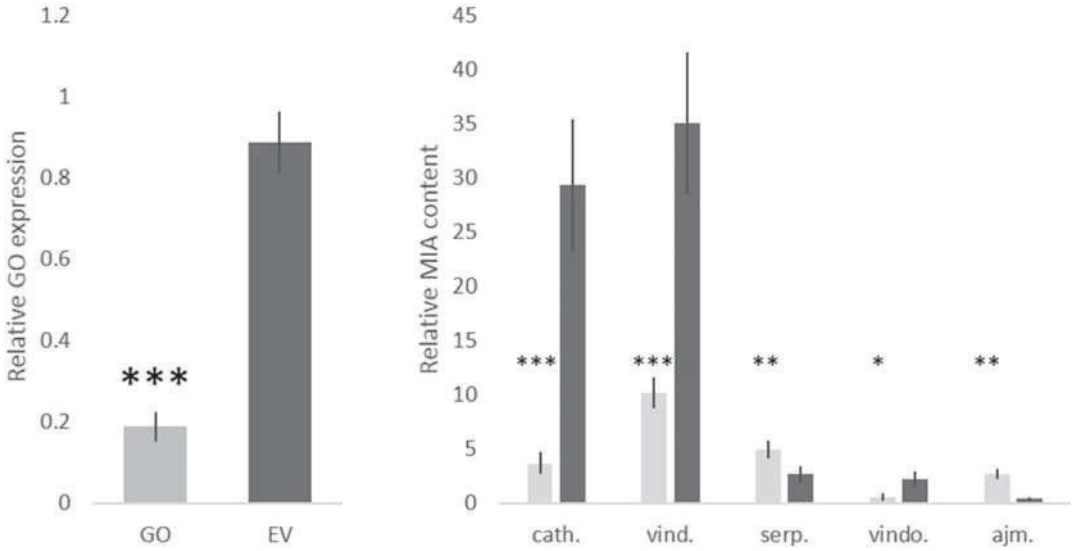


Fig. 5 Biolistic approach using TRV vectors successfully silenced GO in *C. roseus* leaves. (**a**) Relative expression of GO in leaves of pTRV2-EV (grey bar) and pTRV2-GO (white bar)-transformed plants, measured by qPCR. (**b**) Relative mean alkaloid content in leaves of pTRV2-EV (grey bar) and pTRV2-GO (white bar) transformed plants. (**c**) Cath, catharanthine; vind, vindoline; serp, serpentine; vindo, vindorosine; ajma, ajmalicine. Data represent mean ± SE of three technical replicates performed on five plants transformed with pTRV2-EV or pTRV2-GO. Asterisks denote statistical significance ($^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$, Student's *t*-test)

6. Spin down the tubes at 15,000 × *g* for 15 min; collect the supernatant (around 500 μL).
7. Perform a 1:10 (v/v) dilution of the supernatant with the water: acetonitrile solution.
8. Spin down the tubes at 15,000 × *g* for 15 min and collect 50 μL. Load the samples in the proper sampler tubes.
9. The analysis of the alkaloid content was carried out by UPLC-MS.
10. The following multistep gradient was used: acetonitrile (B) versus water (A) system containing 0.1% (v/v) formic acid, at a flow rate of 0.1 mL/min, with 10:90:0.1 to 60:40:0.1 over 18 min.
11. For results described in Fig. 5, MS experiments were carried out in positive mode in the selected ion-monitoring mode using *m/z* 337 for catharanthine ($[M-H]^-$, RT = 11.6 min), *m/z* 457 for vindoline ($[M-H]^-$, RT = 15.4 min), *m/z* 349 for serpentine ($[M+H]^+$, RT = 13.01 min), *m/z* 427 for vindorosine ($[M+H]^+$, RT = 15.03 min), and *m/z* 353 for ajmalicine ($[M+H]^+$, RT = 11.7 min). Relative quantification was performed by correcting peak areas by sample masses dry weight.

3.6 Conclusion

The results described in Fig. 5 shows the efficiency of the silencing approach on the GO target gene. The marked downregulation of GO expression caused by VIGS (around 80%) resulted in substantial decrease of catharanthine, vindoline, and vindorosine accumulation but to an increased synthesis of ajmalicine and serpentine. This pattern of MIA accumulation tempted to demonstrate that GO is likely involved in the early steps of the synthesis of the precursor of *Aspidosperma* (tabersonine, vindoline, vindorosine) and *iboga* (catharanthine). As a consequence, the blocking of the *Aspidosperma* and iboga pathways led to a rerouting of the MIA precursors to other MIA biosynthetic branches such as the heteroyohimbine one ensuring ajmalicine and serpentine synthesis. Such a positioning in the pathway was finally confirmed by biochemical assays enabling to establish that GO was the second enzyme after strictosidine glucosidase in the pathway.

4 Notes

1. Spermidine integrity decreases even kept at −20 °C. To avoid compromising the VIGS procedure, a stock solution of 1 M should be kept at −20 °C followed by 0.1 M aliquots that should be renewed every 3–6 months.
2. Typically, for each round of experiment, 60 plantlets are prepared in order to guarantee the maximum homogeneity of the required developmental stages. VIGS assays should always be planted with at least ten plants per construct including empty vector, candidate gene, and reporter gene. Silencing of the reporter gene causes phenotype alterations (photobleaching for PDS) and is used as a positive control for the procedure and to determine the optimal period for sample harvesting. A second round of silencing is required to confirm results.
3. The ratio of absorbance values at 260 and 280 nm allows the prediction of DNA purity. A pure DNA has an A260/A280 ratio of 1.8–2.0; if the ratios are below this average, it indicates the presence of contaminants such as proteins.
4. To achieve optimal RT performance, take in consideration that a 20 μL RT reaction efficiently converts a maximum of 1 μg of total RNA to cDNA.
5. The target genes should be amplified with primers that generate fragments around 200–400 bp. To avoid cross-silencing of nontarget gene(s), only include sequences very specific to the target gene and that can generate any 21 mer sequence identical to any other gene. 5′ or 3′ untranslated region (UTR) of the target genes can also be included. Avoid homopolymeric regions (poly A or poly T). We use the example of the silencing

of geissoschizine oxidase (GO) from *C. roseus*, using the phytoene desaturase (PDS) reporter gene known to cause the leaf bleaching [28, 30]. The silencing of PDS in *R. tetraphylla* will be also described to illustrate the procedure, but other reporter genes may be also considered such as the one encoding the Protoporphyrin IX Magnesium chelatase subunit chlH (ChlH [31]) whose silencing results in very pale yellow-green leaf phenotypes.

6. For more options of restriction sites used to clone in this vector, *see* the multiple cloning site (MCS) of the pTRV2 vector online. The enzymes used are just suggestions.
7. Never clone the reporter gene (phytoene desaturase, PDS on this case) with your targeting gene. The reporter gene is known to interfere with secondary metabolite pathways.
8. Plasmid integrity should be confirmed by analytical electrophoresis gel performed before each VIGS experiments; plasmid degradation may compromise the experiment.
9. For ten bombardments (ten plant transformations), 6.25 mg of gold particles should be prepared.
10. Best results were achieved when vortex speed ranges 1000 rpm of intensity.
11. Once prepared, particles can be stored in −20 °C upon usage. The preparation of the golden particles does not depend on the species used. Prior to use the golden particles should be sonicated for 15 min with intercalated vortexing periods.
12. To achieve better results, short pulses of ultrasounds with sonic bath should be performed before the gold particles manipulation. During the coating step, vortexing while pipetting avoids precipitation of the beads.
13. Attention: do not exceed the volume of 2 μL of each plasmid—concentrate the plasmid solutions if required.
14. The best conditions allow performing bombardment without resulting in marked leaf damages. This method allows up to 90% of successful silencing of candidate genes in both *C. roseus* and *R. tetraphylla*.
15. On this step, the harvest is performed simultaneously for qPCR and MIA quantification. Thus, a previous identification of your tubes is required to avoid mixing of the samples.
16. Perform this step rapidly to avoid RNA degradation.
17. Be sure to always include duplicate "No-template Negative Controls" (NTC) in which Milli-Q water is used instead of cDNA.
18. The most important step to a success qPCR is high-quality DNA preparation. Integrity and purity of DNA template are

essential. Contaminants can interfere with fluorescence detection.

19. To know the best concentration of primers and matrix for a qPCR reaction, a calibration curve should be done in advance. To generate good calibration curves, different concentrations of DNA (we recommend a minimum of five) should be used. Each concentration should be run in triplicate.
20. At the metabolic level, consequences related to target gene silencing can be evaluated by measuring the qualitative and quantitative differences of MIA accumulation between plants silenced for a dedicated targeted gene and plants transformed with empty silencing vectors. Identifying up- or downregulated MIAs give remarkable clues regarding candidate gene function and positioning in the pathway. The method described for MIA extractions is fitted for a wide range of conditions, and species accumulating this type of compounds but specific adaptations may be required depending on the plant [32].

References

1. Le Quesne PW (1999) Alkaloids: biochemistry, ecology, and medicinal applications edited by Margaret F. Roberts (University of London) and Michael Wink (University of Heidelberg). Plenum Press, New York, NY. J Nat Prod 62 (4):664–664. https://doi.org/10.1021/np980259j
2. Hinse C, Stöckigt J (2000) The structure of the ring-opened N beta-propyl-ajmaline (Neo-Gilurytmal) at physiological pH is obviously responsible for its better absorption and bioavailability when compared with ajmaline (Gilurytmal). Pharmazie 55(7):531–532. PMID: 10944783
3. Ro D, Paradise EM, Ouellet M et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440(7086):940–943. https://doi.org/10.1038/nature04640
4. Song MC, Kim EJ, Kim E et al (2014) Microbial biosynthesis of medicinally important plant secondary metabolites. Nat Prod Rep 31:1497–1509. https://doi.org/10.1039/c4np00057a
5. Galanie S, Thodey K, Trenchard IJ et al (2015) Complete biosynthesis of opioids in yeast. Science 349(6252):1095–1100. https://doi.org/10.1126/science.aac9373
6. Saleki R, Young PG, Lefebvre DD (1993) Mutants of *Arabidopsis thaliana* capable of germination under saline conditions. Plant Physiol 101:839–845. https://doi.org/10.1104/pp.101.3.839
7. Koiwa H, Bressan RA, Hasegawa PM (2006) Identification of plant stressresponsive determinants in Arabidopsis by large-scale forward genetic screens. J Exp Bot 57:1119–1128. https://doi.org/10.1093/jxb/erj093
8. Qu Y, Easson MEAM, Simionescu R et al (2018) Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19E-geissoschizine. Proc Natl Acad Sci U S A 115(12):3180–3185. https://doi.org/10.1073/pnas.1719979115
9. Williams D, Qu Y, Simionescu R et al (2019) The assembly of (+)-vincadifformine and (−)-tabersonine derived monoterpenoid indole alkaloids in *Catharanthus roseus* involve separate branch pathways. Plant J 99 (4):626–636. https://doi.org/10.1111/tpj.14346
10. Guo Y, Halfter U, Ishitani M et al (2001) Molecular characterization of functional domains in the protein kinase SOS2 that is required for plant salt tolerance. Plant Cell 13:1383–1399. https://doi.org/10.1105/tpc.13.6.1383
11. Burgos-Rivera B, Dawe RK (2012) An Arabidopsis tissue-specific RNAi method for studying genes essential to mitosis. PLoS One 7(12):

e51388. https://doi.org/10.1371/journal.pone.0051388

12. Ramegowda V, Mysore KS, Senthil-Kumar M (2014) Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. Front Plant Sci 5:323. https://doi.org/10.3389/fpls.2014.00323
13. Unver T, Budak H (2009) Virus-induced gene silencing, a post transcriptional gene silencing method. Int J Plant Genomics 2009:198680. https://doi.org/10.1155/2009/198680
14. Voinnet O (2001) RNA silencing as a plant immune system against viruses. Trends Genet 17(8):449–459. https://doi.org/10.1016/S0168-9525(01)02367-8
15. Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol 6 (22):3343–3353. https://doi.org/10.1111/j.1365-2958.1992.tb02202.x
16. Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol 2:109–113. https://doi.org/10.1016/S1369-5266(99)80022-3
17. Liu H, Reavy B, Swanson M et al (2002a) Functional replacement of the Tobacco rattle virus cysteine-rich protein by pathogenicity proteins from unrelated plant viruses. Virology 298:232–239. https://doi.org/10.1006/viro.2002.1421
18. Xu P, Zhang Y, Kang L et al (2006) Computational estimation and experimental verification of off-target silencing during posttranscriptional gene silencing in plants. Plant Physiol 142:429–440. https://doi.org/10.1104/pp.106.083295
19. Valentine T, Shaw J, Blok VC et al (2004) Efficient virus-induced gene silencing in roots using a modified Tobacco rattle virus vector. Plant Physiol 136:3999–4009. https://doi.org/10.1104/pp.104.051466
20. Ratcliff F, Martin-Hernandez AM, Baulcombe DC (2001) Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. Plant J 25:237–245. https://doi.org/10.1046/j.0960-7412.2000.00942.x
21. Liu Y, Schiff M, Dinesh-Kumar SP (2002b) Virus-induced gene silencing in tomato. Plant J 31:777–786. https://doi.org/10.1046/j.1365-313X.2002.01394.x
22. Carqueijeiro I, Masini E, Foureau E et al (2015) Virus-induced gene silencing in *Catharanthus roseus* by biolistic inoculation of tobacco rattle virus vectors. Plant Biol 17:1242–1246. https://doi.org/10.1111/plb.12380
23. Corbin C, Lafontaine F, Sepúlveda LJ et al (2017) Virus-induced gene silencing in Rauwolfia species. Protoplasma 254:1813. https://doi.org/10.1007/s00709-017-1079-y
24. Dessens JT, Lomonossoff GP (1993) Cauliflower mosaic virus 35S promoter-controlled DNA copies of cowpea mosaic virus RNAs are infectious on plants. J Gen Virol 74:889–892. https://doi.org/10.1099/0022-1317-74-5-889
25. Jakab G, Droz E, Brigneti G et al (1997) Infectious in vivo and in vitro transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. J Gen Virol 78(12):3141–3145. https://doi.org/10.1099/0022-1317-78-12-3141
26. Foureau E, Carqueijeiro I, Dugé de Bernonville T et al (2016) Prequels to synthetic biology: from candidate gene identification and validation to enzyme subcellular localization in plant and yeast cells. In: O'Connor S (ed) Synthetic biology and metabolic engineering in plants and microbes. Part B: Metabolism in plants, Methods in enzymology, vol 576. Elsevier, Academic Press, Amsterdam, pp 167–206. https://doi.org/10.1016/bs.mie.2016.02.013
27. Lim HS, Vaira AM, Domier LL et al (2010) Efficiency of VIGS and gene expression in a novel bipartite potexvirus vector delivery system as a function of strength of TGB1 silencing suppression. Virology 402(1):149–163. https://doi.org/10.1016/j.virol.2010.03.022
28. Tatsis EC, Carqueijeiro I, Dugé de Bernonville T et al (2017) A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate. Nat Commun 8 (1):316. https://doi.org/10.1038/s41467-017-00154-x
29. Caputi L, Franke J, Farrow SC et al (2018) Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. Science 360(6394):1235–1239. https://doi.org/10.1126/science.aat4100

30. Salim V, Yu F, Altarejos J et al (2013) Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. Plant J 76:754–765. https://doi.org/10.1111/tpj.12330
31. Liscombe DK, O'Connor SE (2011) A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. Phytochemistry 72(16):1969–1977. https://doi.org/10.1016/j.phytochem.2011.07.001
32. Besseau S, Kellner F, Lanoue A et al (2013) A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. Plant Physiol 163(4):1792–1803. https://doi.org/10.1104/pp.113.222828

Article 2: A rapid and efficient vacuum-based agroinfiltration protocol for transient gene overexpression in leaves of *Catharanthus roseus*

This second publication will present the OE approach in detail, which is based on the transient expression technique, a known method for the production of particular proteins or to test expression cassette arrangements using plant cells. This procedure takes advantage of the ability of *Agrobacterium tumefaciens* (Agrobacterium) to transfer a portion of its genome, the T-DNA, into the plant host cell (Gelvin, 2003; Tzfira & Citovsky, 2006). Transient expression systems are very useful because they are fast, flexible, unaffected by chromosomal positional effects and can be used in fully differentiated plant tissues (Fischer et al., 1999).

Technically, Agrobacterium-mediated transient expression is based on the infiltration of Agrobacterium cultures into plant leaves. The Agrobacterium will subsequently transfer a DNA segment, the transfer-DNA or T-DNA, into the plant cells. Originally, the T-DNA is part of the bacterial tumour-inducing (Ti) plasmid that carries the genes required for its transfer, which will cause crown-gall disease on species in the *Rosaceae* family. In the genetically engineered Agrobacterium strains, the original T-DNA is deleted, but two essential parts of it, its left- and right border, which are responsible for its transfer, are kept and transferred to another plasmid, resulting in a system called the "binary vector". In this way, any gene that is placed between these borders, will be transferred to the plant (Voinnet et al., 2003; Zheng et al., 2021).

An obstacle for the agroinfiltration approaches is the post-transcriptional gene silencing (PTGS) previously described in the VIGS article. What is an advantage for the VIGS approach, can be an obstacle for the Agrobacteria-mediated transformations. Since in plants, the PTGS system operates as an adaptive immune against viruses (Ratcliff et al., 1999; Voinnet et al., 2000), many plant viruses have evolved proteins that suppress various steps of the mechanism (Voinnet et al., 1999; 2000). In order to overcome this barrier, an alternative to the original binary vectors has been developed: the pEAQ series of vectors, which contains deconstructed viral genes that enhance recombinant protein production, by producing a highly stable mRNA that is very efficiently transcribed. In addition, it also contains an expression cassette for the P19 suppressor of gene silencing within the T-DNA borders (Sainsbury et al., 2009; Sainsbury & Lomonossoff, 2008).

In the next article, I was responsible for the GFP overexpression in *C. roseus* plants following the protocol designed by the authors, in order to illustrate the process and facilitate the description.

RESUME EN FRANÇAIS DE L'ARTICLE 2

Les analyses génomiques fonctionnelles *in planta* peuvent être entravées chez les plantes non modèles qui sont récalcitrantes à la transformation génétique telle que la plante médicinale *Catharanthus roseus* (Apocynaceae). A ce jour, aucune transformation stable suivie d'une régénération des plantules n'a été réalisée avec un rendement élevé pour cette plante. En outre, alors que le silençage génique transitoire à médiation virale chez *C. roseus* a été démontré il y a une décennie, les outils pour la surexpression transitoire restent rares. Ici, nous décrivons une méthode efficace et fiable pour surexprimer de manière transitoire dans les feuilles de *C. roseus* tout gène d'intérêt. Ce protocole combine une approche d'agroinfiltration sous vide et une haute efficacité traductionnelle d'un vecteur binaire à base de virus déconstruit (pEAQ-HT). La méthode décrite est robuste, facile à réaliser et conduit à une forte expression transitoire chez *C. roseus*. Ce protocole devrait servir d'outil précieux pour améliorer la caractérisation *in planta* de la fonction des gènes, voire pour introduire de manière transitoire de nouvelles activités enzymatiques.

A rapid and efficient vacuum-based Agroinfiltration protocol for transient gene overexpression in leaves of *Catharanthus roseus*.

Authors: Konstantinos Koudounas*, Ines Carqueijeiro, Pamela Lemos Cruz, Jennifer Perrin, Arnaud Lanoue, Audrey Oudin, Sébastien Besseau, Vincent Courdavault*

EA2106 Biomolécules et Biotechnologies Végétales, Université de Tours, 37200 Tours, France

***Corresponding Authors**: Vincent Courdavault; vincent.courdavault@univ-tours.fr

Konstantinos Koudounas; koudounas@univ-tours.fr

Running Title: Vacuum-based Agroinfiltration protocol in Catharanthus

Key Words: *Catharanthus roseus*; Madagascar periwinkle; Apocynaceae; *Agrobacterium tumefaciens*; Agroinfiltration; Vacuum infiltration; Transient expression; Functional Genomics; Cowpea mosaic virus HT (CPMV-HT); Recombinant proteins;

Abstract

Functional genomics analyses *in planta* can be hampered in non-model plants that are recalcitrant to genetic transformation such as the medicinal plant *Catharanthus roseus* (Apocynaceae). No stable transformation and regeneration of plantlets have been achieved with a high efficiency in this plant to date. In addition, while virus mediated transient gene silencing has been reported a decade ago in *C. roseus*, tools for transient overexpression remain scarce. Here, we describe an efficient and reliable methodology for transiently overexpressing any Gene Of Interest in *C. roseus* leaves. This protocol combines a vacuum-based agroinfiltration approach and the high translational efficiency of a deconstructed virus-based binary vector (pEAQ-HT). The described methodology is robust, easy to perform and results in high amount of transient expression in *C. roseus*. This protocol is expected to serve as valuable tool to

enhance the *in planta* characterization of gene functions or even transiently knock-in novel enzymatic activities.

1. Introduction

The Madagascar periwinkle (*Catharanthus roseus*; Apocynaceae) is an emblematic medicinal plant biosynthesizing a plethora of bioactive monoterpene indole alkaloids (MIAs). While these MIAs serve as defense compounds *in planta*, some of them are also of importance for human pharmacopoeia such as the anticancer drugs vinblastine and vincristine [1,2]. Over years, researchers from several laboratories have continuously contributed to the elucidation of MIA biosynthetic pathways and to the discovery of the different enzymes composing the MIA biosynthetic machinery. As a non-model plant, functional genomic approaches remained scarce until recently. The characterization of candidate genes/enzymes was thus mostly based on expression in heterologous hosts and was limited to *in vitro* assays, providing limited knowledge about the actual *in planta* role.

Over the last decade, a tremendous boost in functional genomics studies was achieved in *C. roseus* through the transient silencing of candidate genes (usually named Gene of Interest, GOI) *via* a Virus Induced Gene Silencing approach (VIGS). Such a transient silencing was obtained after delivering genetic constructs encompassing the bipartite RNA genome of Tobacco Rattle Virus (TRV) *in planta* either by Agrobacterium-mediated inoculation [3] or by biolistic transformation [4]. VIGS thus enabled the *in planta* validation of gene functions, up to the elucidation of the biosynthetic pathway of the anti-cancer drugs vinblastine and vincristine [5]. Ideally, beside gene silencing, gene function should be also validated and characterized by GOI overexpression *in planta*. However, stable transformation and overexpression of a GOI are hampered in *C. roseus* since this species is recalcitrant to genetic transformation and regeneration protocols [6]. As a consequence, most of the genetic transformation approaches developed so far in *C. roseus* have been performed in hairy roots or cell suspensions [7] whilst these protocols are either heavily time-consuming or need experienced personnel.

On the other hand, only a few successful transient transformations have been reported in *C. roseus* notably in seedlings [8,9], petals [10] but remained limited in leaves [11,12]. The recent advances in plant virology and the identification of specific sequences enhancing the

replication and/or expression within the plant cells, have resulted in the development of novel deconstructed virus-based vectors that enable transient overexpression at levels overtaking stable transformation approaches [13]. One prominent example is the pEAQ-HT vector which encompasses the non-replicating Cowpea Mosaic Virus-based *HyperTrans* (CPMV-HT) system comprised of a modified 5'-untranslated region (UTR) and the 3'-UTR from CPMV RNA-2 under the control of the Cauliflower mosaic virus (CaMV) 35S promoter and the nopaline synthase (nos) terminator (**Fig. 1**). Additionally an expression cassette for the P19 suppressor of gene silencing has been introduced within the T-DNA borders resulting in high levels of transient expression [14] allowing the plasmid to be used in several studies [15]. It is worth noting that an autonomously replicating viral expression system based on *Catharanthus* yellow mosaic virus (CaYMV) has been lately reported [16].

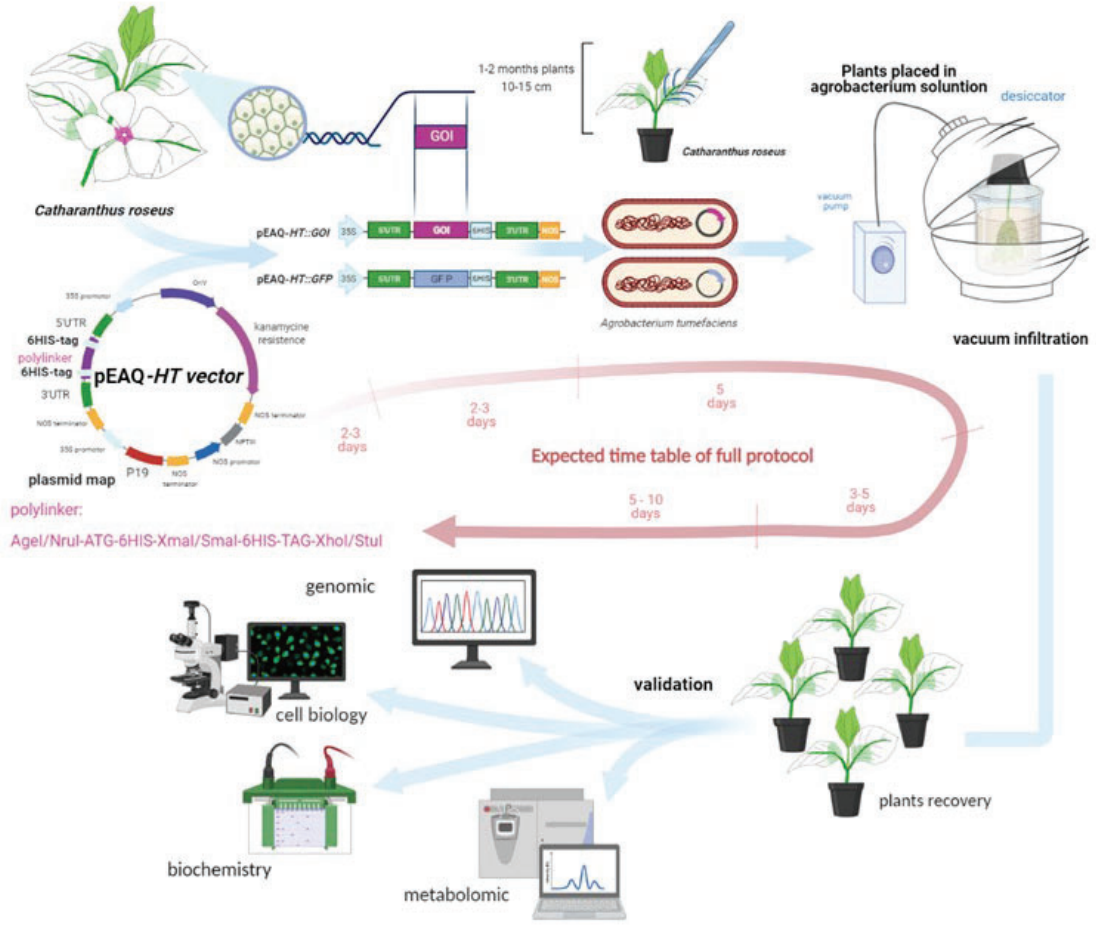


Fig. 1. Schematic overview of the vacuum-based Agroinfiltration protocol in *C. roseus* plants. A GOI is introduced in the polylinker site of the pEAQ-HT vector encompassing the modified 5' and 3' UTRs of CPMV resulting in increased translation of GOI. The plasmid is transformed in *A. tumefaciens* and delivered to *C. roseus* plants by vacuum-infiltration. After recovery,

samples can be collected and used for microscopic, metabolic, genomic or biochemical analyses.

We have recently developed an optimized vacuum-based agroinfiltration protocol for *C. roseus* plants resulting in high levels of transient expression in aerial parts of the plant by recruiting the pEAQ-HT vector (**Fig. 1**). This methodology has already been successfully applied in order to overexpress candidate genes in *C. roseus* [17] and is expected to boost functional genomics studies complementing the VIGS-based approach. Additionally, this approach sets the bases to initiate *in planta* metabolic engineering of *C. roseus* and explore novel chemotypes. Through transient expression of enzymes from distinct species and using the endogenous pool of MIA biosynthetic intermediates as leverage, production of tailor-made or/and new-to-nature MIAs can be investigated. In this chapter, we described in a detailed manner the critical steps of our transient transformation protocol to allow the *C. roseus* community to perform a controlled transient overexpression of genes *in planta*.

2. Materials

2.1. Vectors Construction

1. Sterilized tips, 0.2 and 1.5 mL tubes.
2. Agarose gel apparatus (gel tank, gel tray, well combs and casting tray) and relative equipment (UV transilluminator).
3. 6X Gel loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 40% sucrose).
4. Tris-acetate buffer (TAE). TAE can be prepared as a 50X stock solution by mixing 242 gr Tris-base, 57.1 mL glacial acetic acid and 100 ml of 0.5M EDTA (pH 8.0) in one Liter of ddH_2O.
5. Powdered agarose.
6. 10 mg/mL Ethidium bromide (stored in room temperature and protected from light). Working solution is 0.5 μg/mL.
7. Restriction enzymes (depending on the selected cloning strategy) and the corresponding buffers (see Section 3.1).
8. Alkaline Phosphatase.
9. Heat blocks or water baths.
10. pEAQ-HT vector (GenBank Acc. nr.: GQ497234) [14].

11. pEAQ-HT:GFP-6HIS vector.
12. Kit for RNA extraction.
13. Spectrophotometer.
14. DNase I treatment kit and RNase inhibitor.
15. cDNA synthesis kit including oligo(dT)$_{17}$.
16. Proofreading DNA polymerase and buffer.
17. 10 mM dNTPs set.
18. GOI-specific primers including desired restriction sites (depending on the selected cloning strategy - see Section 3.1).
19. PCR machine.
20. T4 DNA ligase.
21. PCR purification and miniprep kits.
22. Chemically competent *Escherichia coli* (Top10) cells.
23. Luria-Bertani broth medium: 1% tryptone, 0.5% yeast extract and 1% NaCl. In case of solid medium, agar is added in a final concentration of 1.4% before autoclaving.
24. 50 g/L Kanamycin solution prepared in ddH_2O and stored at -20°C (1000X stock solution)).
25. Orbital shakers
26. Centrifuges
27. Incubators at 37 °C.

2.2. Handling of *Agrobacterium* cells

1. *Agrobacterium tumefaciens* strain C58C1 Rif^R (GV3101) containing the T-DNA deficient Ti plasmid pMP90.
2. LB medium.
3. Antibiotic stock solutions: 5 g/L rifampicin resuspended in methanol (100X), 50 g/L gentamicin resuspended in ddH_2O (1000X), 50 g/L kanamycin resuspended in ddH_2O (1000X). All antibiotic stocks are stored at -20 °C. Working solution for each antibiotic is 50 mg/L.
4. Electroporator and electroporation cuvettes of a 0.1 cm gap.
5. Refrigerated centrifuge, centrifugation bottles and tubes.
6. Erlenmeyer flasks, shakers and incubators at 28 °C.

7. Dilution Buffer: prepare stock solutions of 100 mM 2-N-morpholino-ethanesulfonic acid (MES) pH 5.6 and 100 mM $MgCl_2$. Prepare fresh Dilution Buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$) prior agroinfiltration.
8. 100 mM acetosyringone stock solution in dimethyl sulfoxide (DMSO).
9. Sterile glycerol (100%).

2.3. Preparation of *Catharanthus roseus* plants

1. *Catharanthus roseus* seeds (cv. Little Bright Eyes).
2. Soil composed of a 4:1 volume ratio of loam and vermiculite.
3. Greenhouse under controlled conditions (28 °C, 50% relative humidity, 16/8 h light/dark cycle and 70 μmol m^{-2} s^{-1} light intensity).
4. Small trays for seedling cultivation.
5. Square pots (9 cm).
6. Boxes to contain small trays that can be covered with transparent plastic domes to achieve high relative humidity.
7. Razor blades or scalpels.
8. Piece of cardboard or plastic wrap to hold the soil in the pot.

2.4. Vacuum Equipment

1. Vacuum pumping unit.
2. Vacuum desiccator unit of 25cm inner diameter.
3. Tall-form beakers of 150 mL volume capacity.

2.5. Macro/Microscopic analysis

1. UV lamp for macroscopic observation of GFP.
2. Epifluorescence microscope equipped with a proper light source (mercury short arc HBO or LED) and a specific filter set for GFP fluorescence observation (e.g., 460–480 nm excitation filter, 500–520 nm band pass emission filter). The microscope should be equipped with a digital camera and a proper imaging software.
3. Coverslips (24 x 50 mm) and microscope slides (76 x 26 mm).

2.6. Extraction of Proteins and Western Blot

1. Liquid nitrogen
2. Mortar and pestle.
3. Protein Extraction Buffer (100 mM Tris-HCl, pH 8.8, 100 mM EDTA, 1mM PMSF, 100 mM KCL, 100 mM glycine).
4. Heat block.

5.
6. 30% Acrylamide/Bis-acrylamide solution (29:1).
7. 1.5 M Tris-HCL, pH 8.8.
8. 0.5 M Tris-HCL, pH 6.8.
9. 10% Sodium dodecyl sulfate (SDS).
10. 10% ammonium persulfate (APS). Prepare in ddH_2O and store aliquots of 200 µL in -20 °C. Do not thaw and refreeze an aliquot.
11. *N*,*N*,*N*′,*N*′-tetramethylethylene diamine (TEMED).
12. SDS-PAGE Running Buffer: 1X Tris-glycine electrophoresis buffer (25 mM Tris base, 250 mM glycine, 0.1% SDS) can be prepared as 5X or 10X stock solutions.
13. 4X SDS gel-loading loading buffer (200mM Tris-HCL, pH 6.8, 8% SDS, 40% glycerol, 0.4% bromophenol blue, 5% β-mercaptoethanol).
14. Protein electrophoresis system.
15. CBB Staining Solution (45.5% methanol, 9.2% acetic acid, 0.125% Coomassie Brilliant Blue R-250).
16. CBB Destaining Solution (5% methanol, 7.5% acetic acid, 2% glycerol).
17. Bradford protein assay kit.
18. Prestained Protein Ladder.
19. Blotter for semidry transfer.
20. Western blot boxes.
21. Polyvinylidene fluoride (PVDF) transfer membranes.
22. Blotting filter paper.
23. Methanol.
24. 10X Transfer Buffer (480 mM Tris base, 390 mM glycine, 0.375% SDS). Prepare working solution (1X) including 20% methanol.
25. TBST buffer (10mM Tris-HCL pH 8, 150mM NaCl, 0.1% Tween-20).
26. Blocking Solution (5% non-fat dry milk in TBST). Store at – 20 °C. Can be re-used several times.
27. Antibodies: dAnti-6xHis tag primary antibody and a horseradish peroxidase-conjugated secondary antibody, prepared in blocking solution.
28. Orbital shaker.
29. Horseradish Peroxidase (HRP) chemiluminescent substrate.
30. Imaging system for chemiluminescence detection.

3. Methods

3.1 Construction of vectors for transient expression

In this protocol, the pEAQ-HT vector [14] is used for transient overexpression in *C. roseus* plants. This plasmid enables the expression of either N- or C- terminal 6xHIS tagged proteins depending on the polylinker sites that the insert will be introduced (see **Note 1**).

1. Extract RNA from a appropriated source (i.e. plant tissue) using a RNA extraction kit, treat the sample with DNase following the manufacturer's instructions and determine the quantity and the quality of the extracted DNA-free RNAs using a spectrophotometer and agarose-gel electrophoresis, respectively.
2. Perform cDNA synthesis using a commercial kit and following the manufacturer's instructions.
3. Amplify the GOI by PCR on cDNA with a proofreading polymerase and a set of primers that introduce specific restriction sites at both extremities according the selected cloning strategy (i.e. N- or C-terminal tagged protein) (see **Note 1**).
4. Verify by agarose-gel electrophoresis that the amplification resulted in a single band of the expected size.
5. Purify the PCR product by a commercial kit and digest both the product and the pEAQ-HT plasmid with the respective restriction enzymes following the manufacturer's instructions. Perform the double digestions sequentially if needed and heat-inactivate the enzymes.
6. Dephosphorylate the linearized vector with an alkaline phosphatase following the manufacturer's instructions and purify the digestions using a commercial kit.
7. Determine the concentrations of the digested insert and vector by a spectrophotometer.
8. Use a commercial T4 ligase to ligate the insert and the vector following the manufacturer's instructions.
9. Transform the reaction in chemically competent *E. coli* cells by a heat-shock transformation protocol.

10. Plate the cell suspension in LB plates supplemented with 50 mg/L kanamycin, grow overnight at 37 °C and screen the colonies by colony-PCR for the presence of the insert.
11. Inoculate the positive clones in 5 mL liquid LB medium supplemented with 50 mg/L kanamycin, grow at 37 °C under continuous shaking (200-250 rpm).
12. Use a commercial kit to extract the plasmids, validate the presence of the insert after digestion with different combinations of restriction enzymes and verify the constructs by Sanger sequencing. Prepare glycerol stocks for archieving the strains at -80 ℃.

3.2 Agroinfiltration of *C. roseus* plants

3.2.1 Preparation of *C. roseus* plants

1. Germinate the seeds of *C. roseus* (cv. Little Bright Eyes) in soil composed of a 4:1 volume ratio of loam and vermiculite at 28 °C in the greenhouse.
2. When the cotyledons are fully expanded, transfer the plantlets in individual pots and cultivate under the same controlled conditions.
3. Use plants 1-2 months old for vacuum infiltration with 3-5 fully developed pair of leaves.

3.2.2 Preparation of Agrobacterium strains for Agroinfiltration

The *Agrobacterium tumefaciens* GV3101 (pMP90) strain is used in this methodology. Follow standard protocols [18] for preparing electrocompetent Agrobacterium cells and electroporating the strain with the pEAQ-HT:GFP-6HIS or pEAQ-HT:GOI-6HIS plasmids.

1. Streak the GV3101 (pMP90) *Agrobacterium* strains transformed with the pEAQ-HT:GFP-6HIS or pEAQ-HT:GOI-6HIS plasmids in LB plates supplemented with 50 mg/L rifampicin, 50 mg/L gentamicin and 50 mg/L kanamycin and grow for 2 days at 28 °C.
2. Inoculate single colonies from each strain in 3 mL of liquid LB medium supplemented with 50 mg/L rifampicin, 50 mg/L gentamicin and 50 mg/L kanamycin and grow at 28 °C for ~24 h under continuous shaking (200-250 rpm).
3. Initiate 50 mL cultures from each strain in liquid LB supplemented with 50 mg/L rifampicin, 50 mg/L gentamicin and 50 mg/L kanamycin by diluting

1/250 and grow overnight (~14-16hrs) under continuous shaking (200-250 rpm).

4. Centrifuge the *Agrobacterium* cultures for 10 min at 1000g in room temperature (**Fig. 2A**).
5. Wash each pellet with 20 mL of Dilution Buffer without disrupting the pellet in order to remove any remaining LB medium.
6. Gently resuspend the pellet of each culture in 20 mL Dilution Buffer.
7. Measure the OD_{600} of a 1/10 dilution with a spectrophotometer.
8. Dilute the cell suspensions in Dilution Buffer and combine GOI and GFP strains so that each has a final OD_{600}=0.25 (see **Note 2**) in a final volume of 150 mL supplemented with acetosyringone to a final concentration of 150 µM. Use Erlenmeyer flasks (**Fig. 2B**).
9. Incubate the cell suspensions at 28 °C for 3 h under slow agitation.

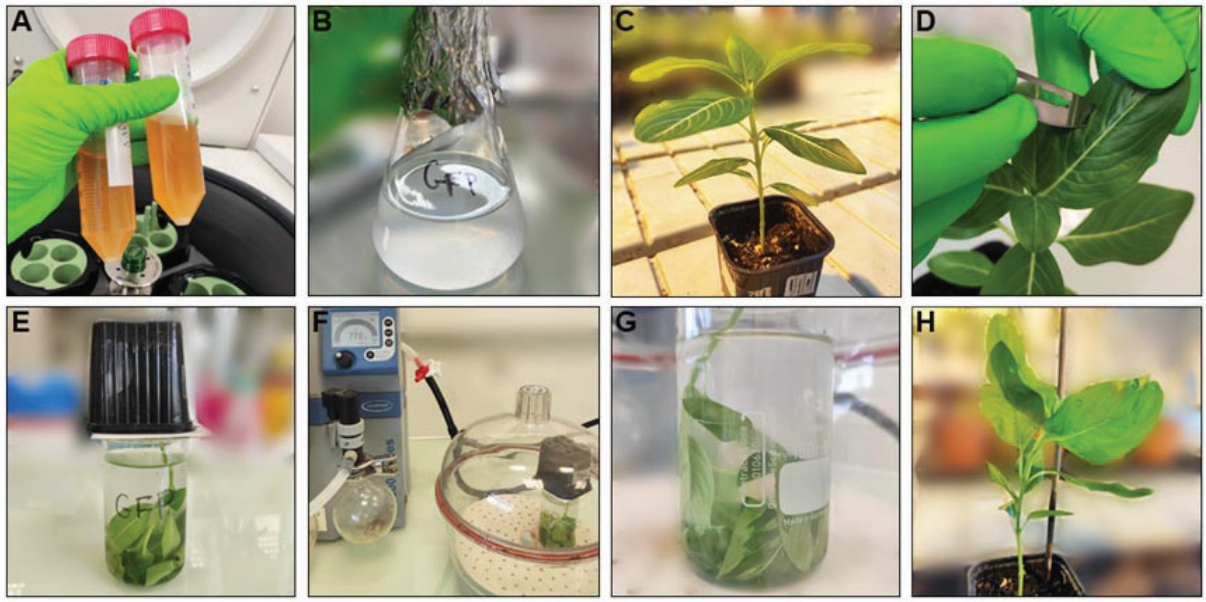


Fig. 2. Vaccum-based Agroinfiltration procedure of *C. roseus* plants. Cultures of *A. tumefaciens* cells are centrifuged (**A**), resuspended in Dilution Buffer at a final OD_{600}=0.25 and incubated at 28 °C for 3 hrs with acetosyringone (**B**). *C. roseus* plants with 3-5 fully developed pair of leaves (**C**) are carefully traumatized with a razor blade between the leaf veins (**D**) and fully submerged in Agrobacterium cell suspensions (**E**). The desiccator is sealed and vacuum is applied (**F**). After restoring the atmospheric pressure, darkened leaf areas appear indicating a successful agroginfiltration (**G**). Five rounds of vacuum infiltration ensure that leaves are homogenously infiltrated. Plants are recovered in the greenhouse under controlled conditions (**H**).

3.2.2 Vacuum infiltration

1. Select *C. roseus* plants with 3-5 fully developed pair of leaves (i.e. 1-2 months old) (**Fig. 2C**).
2. Expose the plants to high humidity conditions 16-24 hours before agroinfiltration. Use a box that can be covered with a transparent plastic dome, add water in the bottom and put the tray with the plants inside the box so that the tray floats (see **Note 3**).
3. Gently traumatize the fully developed leaves by scarifying with a razor blade or a scalpel. Apply 8-10 cuttings of 0.5 cm evenly distributed to a fully developed leaf. Traumatize only in leaf areas between the veins of the leaf (**Fig. 2D**).
4. Transfer the *Agrobacterium* cell suspensions in 150 mL tall-form beakers (**Fig. 2E**).
5. Invert the pot with a *C. roseus* plant and submerge the aerial parts into the beaker containing the *Agrobacterium* cell suspensions (**Fig. 2E**). Ensure that all the leaves are submerged in the solution. Use a piece of cardboard or plastic wrap to retain the soil in the pot during vacuum infiltration.
6. Place the beaker in the center of the vacuum desiccator unit, and properly seal the desiccator unit (**Fig. 2F**).
7. Initiate the vacuum pump and apply a negative pressure of 50 mbar, retain this pressure for 1 minute and gently break the vacuum until atmospheric pressure is reached. Repeat this step for a total of five rounds to ensure that leaves are evenly infiltrated. Typically, after the first or second round, darkening of leaf areas can be observed which indicates a successful infiltration (**Fig. 2G**).
8. Remove the pots from the beakers (**Fig. 2H**), place the infiltrated plants in a tray and return the plants to the greenhouse. Ensure sufficient watering every 2-3 days.

3.3 Validation of transient expression

3.3.1 Macroscopic observation of GFP fluorescence.

1. Observe the infiltrated *C. roseus* plants for GFP fluorescence with a UV lamp at 3 days post infiltration (see **Note 4**) in a dark room.

2. Collect two or three representative samples (i.e. by cutting regions of approx. 1 cm and avoiding the veins) for confirmation of GFP expression under an epifluorescence microscope (see part 3.3.2.).
3. Collect leaves (or leaf areas) exhibiting high fluorescence (**Fig. 3A**) and freeze immediately in liquid nitrogen (see **Note 5**).
4. Store the samples at -80 ℃ until needed.

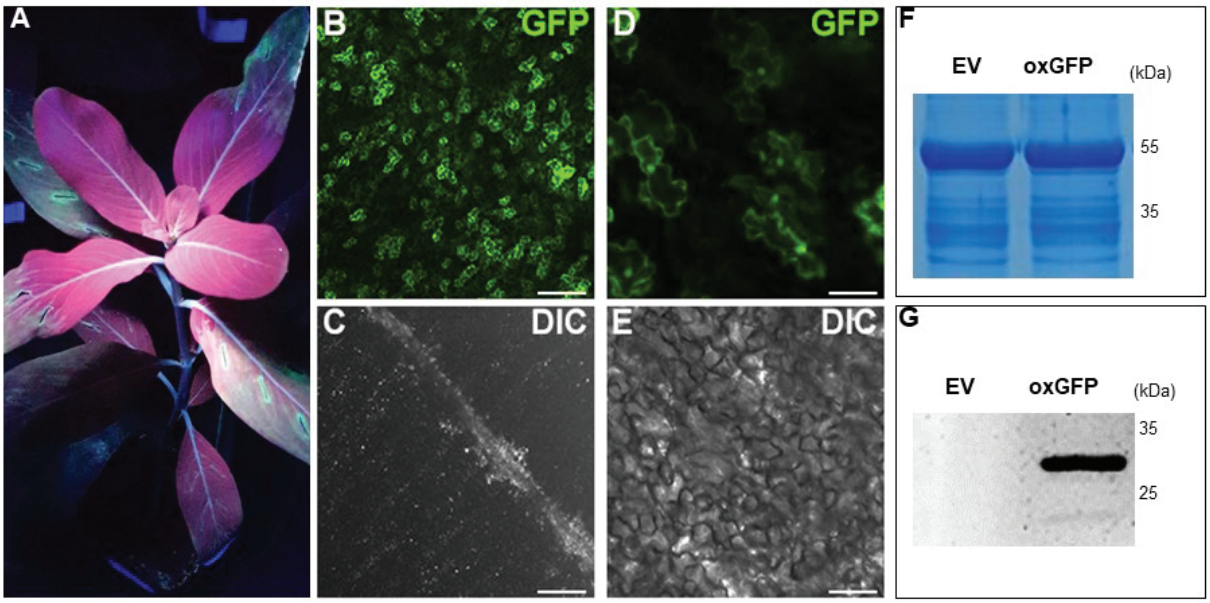


Fig. 3. Validation of transient overexpression of GFP in *C. roseus* leaves. Plants are examined macroscopically in a dark room under a UV lamp at 3 dpi and leaf areas exhibiting GFP fluorescence are collected (**A**). Overexpression of GFP is further confirmed under an epifluorescence microscope (**B-E**) and by immunodetection in soluble protein extracts from the leaf tissues (**F, GF**). Bars 20 µm (**B**, **C**) or 100 µm (**D**, **E**), DIC, differential interference contrast. Protein extracts (20 µg) of *C. roseus* leaves agroinfiltrated with either pEAQ-HT (EV) or pEAQ-HT:GFP-6HIS (oxGFP) constructs were resolved in 10% polyacrylamide gels and bands were visualized after CBB staining (**F**) or Western blotting (**G**) after immunodetection of the 6xHIS epitope. Numbers on the right (**F**, **G**) indicate molecular weights in kDa.

3.3.2 Microscopic observation of GFP fluorescence.

1. Deposit the samples collected for microscopic observation on a microscope slide. The abaxial side of the leaf sample should face to the microscope objective lens.
2. Add a drop of distilled water and shield with a coverslip.

3. Observe the slide under an epifluorescence microscope and acquire images with different objective lenses to ensure that a homogenous expression of GFP was achieved (**Fig. 3B-E**).

3.3.3 Protein extraction.

1. Pre-chill a mortar and a pestle with liquid nitrogen.
2. Transfer the harvested infiltrated *C. roseus* leaves from -80 ℃ to the cold mortar and grind the samples into a fine powder in the presence of liquid nitrogen. Supplement liquid nitrogen periodically to avoid thawing.
3. Transfer 500 mg of the grinded sample in a sterile pre-chilled microcentrifuge tube and immediately add 500 µL of pProtein eExtraction bBuffer. Keep the remaining sample at -80 ℃ in case needed.
4. Mix thoroughly by vortexing the sample for 2 min.
5. Centrifuge the sample at 13.000 g for 10 min at 4 ℃, collect the supernatant and transfer in a sterile pre-chilled microcentrifuge tube. Keep the sample on ice.
6. Determine the concentration of the extracted soluble protein by a Bradford protein assay kit and a spectrophotometer following the manufacturer's instructions. Proceed with SDS-PAGE analysis or store the sample in -20 ℃ until needed.

3.3.4 SDS-PAGE.

1. Assemble the glass plates according to the manufacturer's instructions. Prepare two identical gels, one for Coomassie staining and one for Western blotting.
2. Prepare 10 mL of 10% polyacrylamide resolving gel (see **Note 6**). Mix 4 mL of ddH_2O, 3.3 mL of Acrylamide/Bis-acrylamide solution (30%), 2.5 mL of 1.5 M Tris-HCL (pH 8.8), 100 µL of SDS (10%) and 100 µL of APS (10%) (see **Note 7**). Add 5 µL of TEMED, mix gently and immediately pour the acrylamide solution into the gap between the glass plates of the vertical cassette provided with 0.75 mm thick spacers. Leave enough space (~2.5 cm) for the stacking gel. Gently overlay the gel with ddH_2O.
3. After polymerization is complete, pour off the overlay from the cassette and use the edge of paper towel to completely remove any remaining ddH_2O.

4. Prepare 5 mL of 5% polyacrylamide stacking gel. Mix 2.8 mL of ddH_2O, 830 µL of Acrylamide/Bis-acrylamide solution (30%), 1.26 mL of 0.5 M Tris-HCL (pH 6.8), 50 µL of SDS (10%) and 50 µL of APS (10%) (see **Note 7**). Add 5 µL of TEMED, mix gently and immediately pour the acrylamide solution onto the surface of the polymerized resolving gel. Insert 10-well combs and avoid trapping air bubbles.
5. After polymerization of the stacking gel is complete, assemble the electrophoresis apparatus according to the manufacturer's instructions. Add 1X SDS-PAGE Running Buffer to both electrode compartments, remove the well combs and use a pipette to wash 2-3 times the wells with Running Buffer in order to remove any unpolymerized acrylamide.
6. Dilute 20 µg of soluble protein extracts in Protein Extraction Buffer up to a total volume of 15 µL. Add 5 µL of 4X SDS gel-loading buffer, mix gently, denature the proteins by incubating at 90 °C for 5 minutes and keep the samples on ice.
7. Load each protein sample in the wells and load Prestained Protein Ladder in an empty well.
8. The electrophoresis is conducted with a constant current of 20 mA.
9. Stop the electrophoresis when the tracking dye front (bromophenol blue) is approximately 0.5 cm above the bottom of the glass plates.
10. Use one gel for Coomassie staining in order to observe the protein bands and ensure the integrity of the protein extracts. Immerse the resolving gel in 30 mL CBB staining solution for 20 minutes under gentle agitation. Remove the solution and soak the gel in an equal volume of CBB Destaining Solution for 20 minutes. Repeat this step three or four times until the bands are visible (**Fig. 3F**).

3.3.5 Western Blot and chemiluminescence detection

1. Use the second gel for Western blotting. Equilibrate the resolving gel in 50 mL of 1X Transfer Buffer for 10 minutes.
2. Cut two extra thick blot papers and the PVDF membrane on the exact dimensions of the resolving gel. Activate the PVDF membrane by rinsing in methanol for ~1 minute and soak the membrane and the blotting papers in 1X Transfer Buffer.

3. Assemble the Western blot sandwich in the blotter for semidry transfer. Place one extra thick blot paper on the anode, then the PVDF membrane, then the resolving gel and finally the second extra thick blot paper. Roll out any air bubbles with a 5 mL pipette tip and close the blotter with the cathode and the lid.
4. Connect the blotter to the power supply and apply a constant voltage of 10 V for 60 minutes (see **Note 8)**.
5. After the blotting is finished, disassemble the sandwich and confirm that the prestained protein ladder was evenly migrated to the PVDF membrane. Immediately transfer the membrane into a western blot box containing 25 mL of Blocking Solution. The side of the membrane that was in contact with the resolving gel should be facing up. Incubate the membrane in Blocking Solution for 60 minutes by gently agitating at room temperature. The membrane should be fully submerged in all steps to avoid drying.
6. Discard the Blocking Solution and incubate the membrane in the primary antibody solution for 90 minutes by gently agitating at room temperature (see **Note 9**).
7. Remove the primary antibody solution and wash the PVDF membrane with TBST for 3x10 minutes. The primary antibody solution can be stored at -20 ℃ and reused several times.
8. Incubate the membrane in secondary antibody solution for 45 minutes by gently agitating at room temperature (see **Note 10**). Remove the secondary antibody solution and wash the PVDF membrane with TBST for 3x10 minutes.
9. Prepare the working solution of the substrate of HRP for chemiluminescence detection in a dark room. Mix (1:1) the two components of the substrate in a total volume of 1.5 mL.
10. Take the membrane out of the TBTS solution and remove the excess of TBST by vertically touching one edge of the membrane on a paper towel.
11. Place the membrane on a glass plate, distribute evenly the HRP substrate on the membrane and incubate for 2 minutes in room temperature. Remove the HRP substrate by vertically touching one edge of the plate on a paper towel.

12. Transfer the plate with the membrane in the imaging system for chemiluminescence detection and acquire images at different times of exposure until the immunodetected bands are visible (**Fig. 3G**).

4 Notes

1. In order to generate a N-terminal 6xHIS tagged protein, amplify the GOI with a forward primer which introduces a *Xma*I or *Sma*I restriction site and a reverse primer that introduces a *Xho*I or *Stu*I restriction site and ligate in the respective sites of the plasmid's polylinker. In order to generate a C-terminal 6xHIS tagged protein, amplify the GOI with a forward primer which introduces a *Nru*I or *Age*I restriction site and a reverse primer (without a stop codon) which introduces a *Xma*I or *Sma*I restriction site and ligate in the respective sites of the plasmid's polylinker. If a non 6xHIS tagged protein is desired, amplify the GOI with a forward primer which introduces a *Nru*I or AgeI restriction site and a reverse primer which introduces a *Xho*I or *Stu*I restriction site and ligate in the respective sites of the plasmid's polylinker. *Nru*I, *Sma*I and *Stu*I are restriction enzymes that generate blunt ends. *Age*I and *Xma*I are restriction enzymes that generate cohesive ends which are compatible. Whatever the cloning strategy (blunt-end or cohesive end) used, it is important to dephosphorylate the linearized vector to reduce re-cyclization of the plasmid during ligation without an insert. Compatibility of the ends though increases the options of the restriction sites to be incorporated by the primers in case one occurs in the coding sequence of the GOI. Note that the primers should be designed in respect to the frame of the 6xHIS epitope(s) encoded by the pEAQ-HT plasmid.

2. The optimal final OD_{600} of each culture may be adjusted depending on specific experimental needs. Due to high translational efficiency of the pEAQ-HT vector, an OD600 of 0.25 is enough for high expression and co-infiltration of Agrobacterium cells harboring different constructs (i.e., GFP and one or more GOI). Note that if the OD_{600} of the combined Agrobacterium cell suspension is very high, necrotic symptoms may be observed.

3. High humidity promotes stomatal opening and thus *Agrobacterium* cell suspensions can more easily access the mesophyll intercellular spaces elevating the susceptibility of the leaves to agroinfiltration.

4. Peak of transient expression must be validated for each GOI. Typically, highest protein accumulation is observed at 3-4 dpi but depending on the experiment, more days may be needed until a phenotype is observed. In our experience 7 dpi is usually sufficient to observe alterations on the metabolic content of the leaves and immunodetect the GOI product but, if needed, transient expression of the GOI is immunodetected even at 16 dpi [17].

5. Avoid pooling all the leaves together. Leaves should be collected separately or, if needed, pool only leaves of the same node.. The susceptibility to Agroinfiltration, the efficiency and the stability of the transient expression hugely vary as per age of leaves (i.e. leaves harvested at different node in the same plant).

6. The percentage of acrylamide in the resolving gel can be adjusted depending on the size of the protein of interest and the desired resolution. Typically, the percentage of acrylamide in resolving gels ranges from 6 up to 14 %. A 10% Acrylamide/Bis-acrylamide (29:1) gel has a linear range of separation for proteins of 20 up to 80 kDa.

7. Unpolymerized acrylamide is neurotoxic. Wear gloves and handle with care. APS and TEMED should be added immediately before pouring the gels.

8. The time needed for complete transfer depends on the size of the protein and the percentage of acrylamide in the resolving gel.

9. The primary antibody may be either monoclonal or polyclonal, either raised against a specific epitope (i.e. 6xHIS, FLAG or other) or against a specific recombinant protein. The optimum dilution to be prepared needs to be validated in each case. Typically, primary antibodies are diluted in Blocking Solution in a range of 1:500 up to 1:5000. Store at – 20 °C. Primary antibody solution can be used several times.

10. Select a secondary antibody in respect to the primary antibody that will be used. If the primary antibody is produced for example in mouse, the secondary antibody should be an anti-mouse produced in another animal (i.e. goat, rabbit or other). Secondary antibody is diluted in Blocking Solution and typically in a range of 1:1000 up to 1:3000. The optimum dilution to be prepared needs to be validated in each case. Store at – 20 °C. Avoid using the secondary antibody more than 4-5 times.

Acknowledgments:

We acknowledge funding from the ARD2020 and ARD CVL Biopharmaceutical programs of the Région Centre Val de Loire (BioPROPHARM, CatharSIS and ETOPOCentre projects), EU Horizon 2020 research and innovation program (MIAMi project-grant agreement N°814645), and ANR MIACYC (ANR-20-CE43-0010).

References

1. O'Connor SE, Maresh JJ (2006) Chemistry and biology of monoterpene indole alkaloid biosynthesis. Natural product reports 23 (4):532-547. doi:10.1039/B512615K

2. Dugé de Bernonville T, Clastre M, Besseau S, Oudin A, Burlat V, Glévarec G, Lanoue A, Papon N, Giglioli-Guivarc'h N, St-Pierre B, Courdavault V (2015) Phytochemical genomics of the Madagascar periwinkle: Unravelling the last twists of the alkaloid engine. Phytochemistry 113:9-23. doi:https://doi.org/10.1016/j.phytochem.2014.07.023

3. Liscombe DK, O'Connor SE (2011) A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*. Phytochemistry 72 (16):1969-1977. doi:https://doi.org/10.1016/j.phytochem.2011.07.001

4. Carqueijeiro I, Masini E, Foureau E, Sepúlveda LJ, Marais E, Lanoue A, Besseau S, Papon N, Clastre M, Dugé de Bernonville T, Glévarec G, Atehortùa L, Oudin A, Courdavault V (2015) Virus-induced gene silencing in *Catharanthus roseus* by biolistic inoculation of tobacco rattle virus vectors. Plant Biology 17 (6):1242-1246. doi:https://doi.org/10.1111/plb.12380

5. Caputi L, Franke J, Farrow SC, Chung K, Payne RME, Nguyen T-D, Dang T-TT, Soares Teto Carqueijeiro I, Koudounas K, Dugé de Bernonville T, Ameyaw B, Jones DM, Vieira IJC, Courdavault V, O'Connor SE (2018) Missing enzymes in the biosynthesis of the anticancer

drug vinblastine in Madagascar periwinkle. Science 360 (6394):1235-1239. doi:10.1126/science.aat4100

6. Facchini PJ, De Luca V (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. The Plant Journal 54 (4):763-784. doi:https://doi.org/10.1111/j.1365-313X.2008.03438.x

7. Sharma A, Amin D, Sankaranarayanan A, Arora R, Mathur AK (2020) Present status of *Catharanthus roseus* monoterpenoid indole alkaloids engineering in homo- and hetero-logous systems. Biotechnology Letters 42 (1):11-23. doi:10.1007/s10529-019-02757-4

8. Weaver J, Goklany S, Rizvi N, Cram EJ, Lee-Parsons CWT (2014) Optimizing the transient Fast Agro-mediated Seedling Transformation (FAST) method in *Catharanthus roseus* seedlings. Plant cell reports 33 (1):89-97. doi:10.1007/s00299-013-1514-2

9. Mortensen S, Bernal-Franco D, Cole LF, Sathitloetsakun S, Cram EJ, Lee-Parsons CWT (2019) EASI Transformation: An Efficient Transient Expression Method for Analyzing Gene Function in *Catharanthus roseus* Seedlings. Frontiers in plant science 10 (755). doi:10.3389/fpls.2019.00755

10. Schweizer F, Colinas M, Pollier J, Van Moerkercke A, Vanden Bossche R, de Clercq R, Goossens A (2018) An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in *Catharanthus roseus*. Metabolic Engineering 48:150-162. doi:https://doi.org/10.1016/j.ymben.2018.05.016

11. Di Fiore S, Hoppmann V, Fischer R, Schillberg S (2004) Transient gene expression of recombinant terpenoid indole alkaloid enzymes in *Catharanthus roseus* leaves. Plant Molecular Biology Reporter 22 (1):15-22. doi:10.1007/BF02773344

12. Sharma A, Verma P, Mathur A, Mathur AK (2018) Genetic engineering approach using early Vinca alkaloid biosynthesis genes led to increased tryptamine and terpenoid indole alkaloids biosynthesis in differentiating cultures of *Catharanthus roseus*. Protoplasma 255 (1):425-435. doi:10.1007/s00709-017-1151-7

13. Peyret H, Lomonossoff GP (2015) When plant virology met Agrobacterium: the rise of the deconstructed clones. Plant biotechnology journal 13 (8):1121-1135. doi:10.1111/pbi.12412

14. Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant biotechnology journal 7 (7):682-693. doi:10.1111/j.1467-7652.2009.00434.x

15. Peyret H, Lomonossoff GP (2013) The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant molecular biology 83 (1):51-58. doi:10.1007/s11103-013-0036-1

16. Mortimer C, Dugdale B, Waterhouse P (2020) Development of an autonomously replicating viral expression system tailored for Catharanthus roseus. Plant biotechnology journal 18 (5):1115-1117. doi:https://doi.org/10.1111/pbi.13274
17. Carqueijeiro I, Koudounas K, de Bernonville TD, Sepúlveda LJ, Mosquera A, Bomzan DP, Oudin A, Lanoue A, Besseau S, Cruz PL, Kulagina N, Stander EA, Eymieux S, Burlaud-Gaillard J, Blanchard E, Clastre M, Atehortùa L, St-Pierre B, Giglioli-Guivarc'h N, Papon N, Nagegowda DA, O'Connor SE, Courdavault V (2020) Alternative splicing creates a pseudo-strictosidine β-D-glucosidase modulating alkaloid synthesis in *Catharanthus roseus*. Plant physiology:Article Online. doi:10.1093/plphys/kiaa075
18. Koudounas K, Thomopoulou M, Angeli E, Tsitsekian D, Rigas S, Hatzopoulos P (2020) Virus-Induced Gene Silencing in Olive Tree (Oleaceae). In: Courdavault V, Besseau S (eds) Virus-Induced Gene Silencing in Plants: Methods and Protocols. Springer US, New York, NY, pp 165-182. doi:10.1007/978-1-0716-0751-0_13

Part 2: 16-O-methyltransferase identification and involvement in alkaloids biosynthesis

On this second part of the results, two articles on the identification of 16-O-methryltransferases are presented. The article 3 introduces a new tabersonine-16-*O*-methyltransferase isoform from *Catharanthus roseus*. This isoform plays an important role in the vindoline biosynthesis in the plant and was fully characterized. The study of the different O-methyltransferase isoforms has also made it possible to propose hypotheses regarding their emergence during evolution.

The article 4 presents the identification of an orthologue of the tabersonine-16-*O*-methyltransferase in *Vinca minor*, a small subshrub (chamaephyte) belonging to the family Apocynaceae, also known as lesser periwinkle and dwarf periwinkle. Indole alkaloids are the most important secondary metabolites that have been isolated from this genus. Vinca species are very rich in indole alkaloids. Interestingly, the name Vinca alkaloids has been attributed to the class of "bisindole" alkaloids from *Catharanthus roseus*, and from some other Apocynaceae genera. Until today, only one bisindole alkaloid, vincarubine, has been isolated from a Vinca species (*V. minor*) (Proksa et al., 1988).

Vincamine is the major alkaloid present in aerial organs and it is chemically specific to *V. minor*. This alkaloid is of great pharmaceutical interest due to its modulatory effects on brain circulation and neuronal homeostasis, as well as antihypoxic and neuroprotective properties. Due to its medicinal importance, the functional identification of the metabolic pathway genes involved in vincamine biosynthesis and other MIAs in *V. minor* have been put in place.

The completion of vincristine and vinblastine MIA pathway in *C. roseus* has facilitated access to a vast number of gene libraries allowing the use of bioinformatics tools to elucidate *V. minor* metabolic pathways. In this way, the article 4 presents the use of bioinformatics guided screening with the aim to identify and study candidate genes involved in the aspidosperma pathway from annotated transcriptomic databases of medicinal plants. From bioinformatic screening using *C. roseus* genes as bait, were selected and functionally studied candidate genes homologous to *C. roseus* gene that might encode tabersonine 16-O-methyltransferase-like (16OMT) to finally identify the orthologue Vm16OMT.

Article 3 : Vindoline synthesis is coordinated by two 16-hydroxytabersonine-O-methyltransferase isoforms in leaves of *Catharanthus roseus*

In the article 3, I was responsible for the design and execution of all experiments of gene validation and biochemical assays, being assisted by my superiors. This includes qPCR analyses, molecular cloning, yeast bioconversion, metabolite quantifications, metabolite purification (16-hydroxytabersonine), protein purification, enzymatic assays, kinetic assays, *Tobacco* and *C. roseus* agroinfiltrations, VIGS, overexpression analyses, subcellular localization studies and interaction studies (BifC, protein electrophoresis and western blot). In addition, I was responsible for the statistical analysis in R and for the writing of the paper, with the assistance of my superiors. The transcriptome studies and their analysis were performed by other authors.

RESUME EN FRANÇAIS DE L'ARTICLE 3

La plante médicinale *Catharanthus roseus* biosynthétise de nombreuses molécules dont certaines sont utilisées en santé humaine, notamment les alcaloïdes indoliques monoterpéniques (AIM) vinblastine et vincristine. Au cours des dernières décennies, l'augmentation continue de la demande pharmaceutique a incité plusieurs groupes de recherche à caractériser les voies de biosynthèse des AIM pour envisager de futurs approvisionnements en ayant recours à l'ingénierie métabolique. Conformément aux travaux antérieurs suggérant que la diversification peut potentiellement se produire à différentes étapes le long de la chaîne métabolique conduisant à la vindoline, nous avons étudié l'implication d'isoformes distinctes de la tabersonine-16-O-méthyltransférase (16OMT) qui joue un rôle central dans la biosynthèse des AIM. En combinant des recherches d'homologie basées sur la 16OMT1 précédemment caractérisée, des analyses phylogénétiques, des tests fonctionnels dans la levure et des caractérisations biochimiques, nous avons identifié une seconde isoforme de 16OMT, nommée 16OMT2. 16OMT2 semble être une enzyme multifonctionnelle acceptant à la fois comme substrats des alcaloïdes et des flavonoïdes. Ces résultats suggèrent qu'une évolution contrainte de l'enzyme s'est probablement produite pour favoriser l'apparition de 16OMT2 à partir d'une flavonoïde-O-méthyltransférase spécifique ancestrale. Étant donné que 16OMT1 et 16OMT2 présentent une identité de séquence élevée et des paramètres cinétiques similaires pour la 16-hydroxytabersonine, nous postulons que 16OMT1 peut résulter d'une duplication ultérieure du

gène *16OMT2* accompagnée d'une néofonctionnalisation conduisant à une perte quasi complète de l'activité flavonoïde O-méthyltransférase. Dans l'ensemble, ces résultats participent à une meilleure connaissance des processus évolutifs qui ont probablement conduit à la cooptation enzymatique pour la synthèse des AIM.

Vindoline synthesis is coordinated by two 16-hydroxytabersonine-*O*-methyltransferase isoforms in leaves of *Catharanthus roseus*.

Pamela Lemos Cruz[1], Ines Carqueijeiro[1], Konstantinos Koudounas[1], Dikki Pedenla Bomzan[2], Emily Amor Stander[1], Natalja Kulagina[1], Audrey Oudin[1], Arnaud Lanoue[1], Nathalie Giglioli-Guivarc'h[1], Dinesh A Nagegowda[2], Nicolas Papon[3], Sébastien Besseau[1], Marc Clastre[1], Vincent Courdavault[1*]

[1]*Université de Tours, EA2106 "Biomolécules et Biotechnologies Végétales", Tours, France.*

[2]*Molecular Plant Biology and Biotechnology Lab, CSIR-Central Institute of Medicinal and Aromatic Plants, Research Centre, Bengaluru - 560065, India.*

[3] *Univ Angers, Univ Brest, GEIHP, SFR ICAT, F-49000 Angers, France.*

* Corresponding author: vincent.courdavault@univ-tours.fr (ORCID: 0000-0001-8902-4532)

Abstract

The medicinal plant *Catharanthus roseus* biosynthesizes many important drugs for human health, including the anticancer monoterpene indole alkaloids (MIAs) vinblastine and vincristine. Over the past decades, the continuous increase in pharmaceutical demand has prompted several research groups to characterize MIA biosynthetic pathways for considering future metabolic engineering processes of supply. In line with previous work suggesting that diversification can potentially occur at various steps along the vindoline branch, we were here interested in investigating the involvement of distinct isoforms of tabersonine-16-*O*-methyltransferase (16OMT) which plays a pivotal role in the MIA biosynthetic pathway. By combining homology searches based on the previously characterized 16OMT1, phylogenetic analyses, functional assays in yeast, and biochemical and *in planta* characterizations, we identified a second isoform of 16OMT, referred to as 16OMT2. 16OMT2 appears to be a multifunctional enzyme working on both MIA and flavonoid substrates, suggesting that a constrained evolution of the enzyme for accommodating the MIA substrate has probably occurred to favor the apparition of 16OMT2 from an ancestral specific flavonoid-*O*-methyltransferase. Since 16OMT1 and 16OMT2 displays a high sequence identity and similar kinetic parameters for 16-hydroxytabersonine, we postulate that 16OMT1 may result from a later 16OMT2 gene duplication accompanied by a continuous neofunctionalization leading to an almost complete loss of flavonoid *O*-methyltransferase activity. Overall, these results participate in increasing our knowledge on the evolutionary processes that have likely led to enzyme co-optation for MIA synthesis.

Keywords: alkaloids, *Catharanthus roseus*, O-methyltransferase, gene duplication, neofunctionalization

1. Introduction

Monoterpene indole alkaloids (MIAs) are a group of natural products synthesized to cope with environmental constraints, by several plant families including Nyssaceae from the Cornales order and Rubiaceae, Apocynaceae, Gelsemiaceae and Loganiaceae from the Gentianales order [1]. Among their known properties, MIAs notably exhibit deterrent effects and high toxicity protecting plants from pests and conferring them a central place in the defense arsenal deployed along evolution [2,3]. Such properties have been also exploited in human pharmacopoeia to cure several disorders as illustrated by the antimalarial quinine, the antiarrhythmic ajmaline and the anticancer vinblastine and vincristine [4–6]. These last two MIAs are specifically synthesized in leaves of the Madagascar periwinkle (*Catharanthus roseus*) that are still currently exploited to produce these chemotherapy drugs through hemi-synthesis [7,8]. Over the last 50 years, the continuous increase of vinblastine and vincristine demand has prompted several research groups to characterize MIA biosynthetic pathways up to an almost complete elucidation [9–11] but also more recently, to transfer the corresponding MIA biosynthetic genes in yeast for creating cell factories producing MIAs through metabolic engineering [12–16].

From a general point of view, MIAs result from complex biosynthetic pathways that all start from the common precursor and first MIA strictosidine, except camptothecine in *Camptotheca accuminata* [17,18]. The subsequent decorations, oxydo-reductions and cyclisations of strictosidine give birth to the manifold MIA structures accumulated in a plant-, developmental- and organ-dependent manner [17,19,20]. For instance, vinblastine/vincristine arise from the condensation of catharanthine accumulated in both above and underground *C. roseus* organs, with vindoline, specifically produced in leaves [21–23]. While catharanthine synthesis requires nine steps from strictosidine, the production of vindoline is far more complex and first involves the synthesis of tabersonine whose biosynthetic route only differs from that of catharanthine at the ultimate step [11]. The subsequent tabersonine to vindoline conversion consists of seven enzymatic reactions, which are initiated by tabersonine-16-hydroxylase (T16H; EC 1.14.13.73) [24,25] (**Figure 1**). While this C-16 hydroxylation of tabersonine specifically occurs in aerial parts of the plants, roots host tabersonine-19-hydroxylase initiating the synthesis of hörhammericine via 19-hydroxytabersonine production [14,26–28].

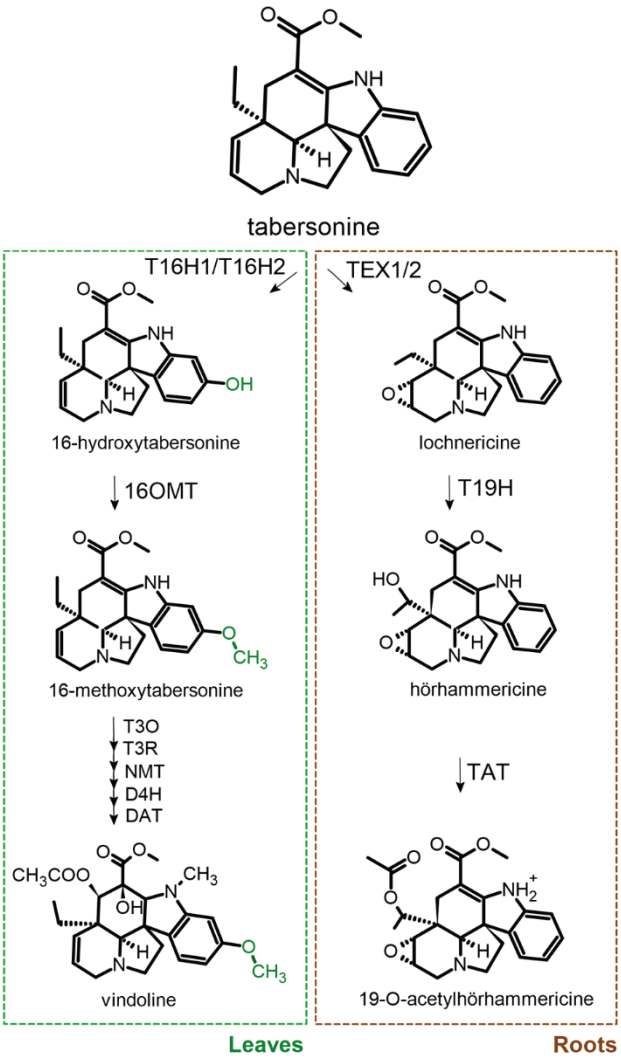


Figure 1. The fate of tabersonine in *C. roseus* is organ-specific. The hydroxylation at the C-16 position is performed by T16H (tabersonine 16-hydroxylase) in leaves and other aerial organs, while in roots, tabersonine is hydroxylated in the position C-19 by T19H (tabersonine 19-hydroxylase). The T16H product is then methylated by 16OMT to produce 16-methoxytabersonine, which after 5 enzymatic steps, is transformed into vindoline. In roots, tabersonine can be epoxidized by tabersonine 6,7-epoxidase isoforms 1 and 2 (TEX1/TEX2), before the C-19 hydroxylation by T19H, resulting in the product hörhammericine, which is acetylated by tabersonine derivative 19-*O*-acetyltransferase (TAT), to produce 19-*O*-acetylhörhammericine. Note that the convention of carbon numbering shown here is not consistent with IUPAC nomenclature but reflects historical versions of numbering found in the enzymatic literature.

By contrast, the T16H-mediated hydroxylation of tabersonine generates 16-hydroxytabersonine, which is then methylated by 16-hydroxytabersonine-16-*O*-methyltransferase (16OMT, 2.1.1.94) [29,30], whose initial identification turned out to be quite

time-consuming. Indeed, while the 16-methyltransferase activity was initially identified in *C. roseus* seedlings [30], several attempts based on homology-based methods failed to clone the corresponding gene but allowed the characterization of several flavonoid-*O*-methyltransferases such as CrOMT2, CrOMT4, CrOMT5 and CrOMT7 [31–33]. Finally, isolation of epidermis combined with protein sequencing and PCR amplification led to the identification of 16OMT [29]. Further studies confirmed the main expression of 16OMT in *C. roseus* leaf epidermis and established that it operates as homodimers in the cytosol [34]. 16OMT produces 16-methoxytabersonine that is then epoxidized by tabersonine-3-oxygenase (T3O) [35], producing a non-stable product rapidly reduced by tabersonine 3-reductase (T3R) to produce 3-hydroxy-16-methoxy-2,3-dihydrotabersonine [13,30]. This is followed by a *N*-methylation catalyzed by 3-hydroxy-16-methoxy-2,3-dihydrotabersonine *N*-methyltransferase (NMT) [36], a hydroxylation performed by desacetoxyvindoline 4-hydroxylase (D4H) [37] and a final acetylation by deacetylvindoline 4-*O*-acetyltransferase (DAT) [30,37,38] to produce vindoline (**Figure 1**).

Besides a marked complexity in terms of enzymatic steps, MIA biosynthetic pathways also exhibit a complex organization at the cellular and subcellular levels with several tissues and organelles hosting each reaction [9] and diversification of the enzymes catalyzing specific MIA decorations. For instance, a pseudoenzyme variant of the strictosidine-β-glucosidase (SGD) ensuring strictosidine activation, has been recently identified as a regulator of MIA biosynthesis [39]. In addition, several enzyme isoforms have been characterized along the MIA pathway notably including two secologanin synthases (SLS) with complementary involvement in the synthesis of the MIA monoterpenoid precursor [10], two tabersonine epoxidases (TEX1; TEX2; [27]) , as well as no less than four isoforms of tetrahydroalstonine synthase (THAS) [40,41]. A prime example of the differential involvement of enzyme isoforms in MIA synthesis is the identification of two T16Hs (T16H1, T16H2) displaying distinct expression profiles with T16H1 expressed in *C. roseus* flowers and T16H2 in leaves. Such a profile thus grants a major role of T16H2 in the total amount of vindoline accumulated in the whole plant [42]. While no other enzyme isoform has been identified to date along the vindoline pathway, diversification can potentially occur at all steps downstream T16H. For instance, different *O*-methyltransferase (OMT) isoforms have been reported to catalyze the same reaction in the synthesis of benzylisoquinoline alkaloid (BIA) in *Papaver somniferum* and *Nelumbo nucifera* [43,44]. More importantly, the formation of OMT heterodimers leads to the synthesis of distinct BIAs, thus increasing the catalytic diversity of OMTs [43]. In *C. roseus*, while several

flavonoid-*O*-methyltransferases have been identified, only one 16OMT has been reported to date [29]. In this work, we described the characterization of a second 16OMT isoform involved in vindoline synthesis but displaying distinct substrate specificities thus shedding light on the evolution process leading to enzyme co-optation for MIA synthesis.

2. Materials and methods

2.1. Chemicals

16-hydroxytabersonine and 19-hydroxytabersonine was obtained by the 16-hydroxylation or 19-hydroxylation of tabersonine by galactose induced yeasts expressing T16H2 or T19H, respectively, as described in [14,42]. Tabersonine was obtained from ChromaDex (Los Angeles, California, USA) and S-adenosyl-L-methionine (SAM) was purchased from Sigma-Aldrich (St-Quentin-Fallavier, France). The flavonoids were obtained from different suppliers, myricetin (Carbosynth), kaempferol (Extrasynthese), naringenin (Sigma-Aldrich) and quercetin (Extrasynthese).

2.2. Isoforms transcript identification

The Gene Model Annotation for *C. roseus* (version 2, [45]) was screened and 21 genes annotated as *O*-methyltransferase, and containing full-length coding sequences were selected for further analysis. A phylogenetic tree was built by blasting (https://blast.ncbi.nlm.nih.gov/Blast.cgi) the 16OMT sequence against the C. roseus genome. Proteins with high percentage of identity were selected. In order to root the tree, the 16OMT sequence was also blasted against the genome of Streptomyces (taxid: 1883) and the protein with the higher percentage of identity was selected. The selected protein sequences were aligned by using the COBALT tool (https://www.ncbi.nlm.nih.gov/tools/cobalt/) and the phylogenetic tree was built on the same tool by using the method “Neighbor Joining” with the parameters: “Max seq difference = 0.85” and the distance Grishin (protein). The final tree figure was designed with iTOL v6 (https://itol.embl.de/).

2.3. Plant and cell culture growth

C. roseus plants were grown at 28 °C under a 16h light/8h dark cycle in a greenhouse. For overexpression experiments, mature plants were used. For VIGS experiments, the plants were grown at 22°C under the same light regime, until the first or second pair of leaves reached full

development [46]. For *Nicotiana benthamiana*, the temperature of growth was 24°C with a photoperiod of 16h/8h light/dark cycle also in a greenhouse. The *C. roseus* CA cell suspension culture used for subcellular localization studies was propagated in Gamborg B5 medium (Duchefa) at 24°C under continuous shaking (100 rpm) for 7 days as described previously [47].

2.4. DNA, RNA Extraction and retro-transcription

Total RNA extractions and DNA purification were performed from distinct *C. roseus* organs (leaves, stems, flowers, flower buds, fruits and roots) using the NucleoSpin RNA Plant and the NucleoSpin Plant II kits (Macherey-Nagel). The epidermis-enriched fraction was obtained from a modified version of a procedure previously described by [48] and modified by [39]. In brief, a cotton swab was dipped into a carborundum powder (particle size <300 grit, Fischer) and used to abrade both lower and upper epidermis layers of young *C. roseus* leaves, which were subsequently dipped in 4 mL of Trizol (Life Technologies) for 5 sec. The RNAs were extracted according to the manufacturer's protocol. The RNA pellet was washed with 70% ethanol and re-suspended in 100 µL of RNAse-free water. Additional cleaning steps were performed by precipitation in ethanol [39]. RNAs from both fractions were quantified using a NanoDrop® ND-1000 and 1 µg were retro-transcribed with the RevertAid First Strand cDNA Synthesis Kit according to the provider's instructions (ThermosFisher Scientific).

2.5. Gene expression analysis

Gene expression measurements were performed using the Dynamo ColorFlash SYBR Green qPCR kit (ThermoFischer Scientific) in a 15 µl final volume containing 6 µl of diluted template cDNA (18 x) and qPCR primers (0.5 µM). Real-time PCR was run on a CFX96 real-time SYBR system (Bio-Rad) using detection of SYBR green with the following conditions: 95°C for 7 min, 40 cycles containing two steps of 95°C for 10 sec and 60°C for 40 sec. Absolute quantification of transcript copy number was performed with calibration curves and normalization with CrRPS9 ratios. All amplifications were performed in triplicate.

2.6. Cloning of candidate isoforms

The coding sequence of the candidate genes CRO_T104105, CRO_T104094 and the gene 16OMT1 were obtained using as matrix the cDNA from *C. roseus* small leaves, with dedicated primers harboring specific restriction sites at their extremities (**Supplemental Table S1)**. The amplified product was cloned into the expression vector pESC_LEU (Stratagene). Due to

technical adversities, CRO_T104094 amplification tailed a different strategy described in (**Supplementary Figure S1**) where the transcripts were obtained following transient expression in *N. benthamiana* leaves. Briefly, the whole sequence of CRO_T104094 was amplified from *C. roseus* genomic DNA, with the specific primers pEAQ_NBCROT104094 (**Supplemental Table 1**) introducing an *Age*I restriction site in both extremities. The recombinant vector was transformed into the GV3101 Agrobacterium strain prior to *N. benthamiana* agroinfiltration as described in [14]. Five days post-infiltration (5 dpi), leaves were collected, and RNA were extracted and retrotranscribed. The generated cDNAs were used to amplify CRO_T104094 using the primers pESC-CRO_T104094.

2.7. Heterologous expression of the isoform candidates in yeast and functional assays

The pESC_LEU plasmids containing the candidate isoforms were independently transformed in to *Saccharomyces cerevisiae* strains WAT11 expressing the *A. thaliana* NADPH P450 reductase 1 [49] together with the T16H2 expression vector (pYeDP60) from [42]. The transformation was followed by the growth of the yeast strains until stationary phase in 4 mL CSM medium (0.67% (*w/v*) yeast nitrogen base, 2% (*w/v*) dextrose, and 0.05% (*w/v*) dropout mix without the respective amino acid for selection, and harvested by centrifugation. Due to the presence of inducible promoters, the protein induction was performed by the inoculation of harvested cells in 10 mL of YPGal medium (2% bacto peptone, 1% yeast extract, and 2% galactose) for 5 h as described by [50]. The yeast constructions used in this study are listed in the **Supplemental Table S2.** The functional assays were performed by feeding the galactose-induced yeast strains with 125 µM tabersonine during 16 h at 30°C under vigorous shaking. The reactions were stopped by centrifugation at 5000 *g* for 10 min. The supernatant fraction was collected and diluted in 3 *v/v* of methanol 100%, vortex and centrifuged at 16,000 *g* for 15 min prior to analyses in ultra-performance liquid chromatography-mass spectrometry (UPLC-MS).

2.8. Production of the recombinant proteins and test of interactions

16OMT1, CRO_T104105, CRO_T104094 and FlavoOMT coding sequences were obtained as described previously, using the specific primers designed to introduce a *Bam*HI restriction site at each extremity (**Supplemental Table S1)** and ligated in pRSET-A plasmids (Thermo Fisher Scientific). Recombinant proteins were produced in *E. coli* BL21 DE3 cells, in all cases, cells were cultivated at 37°C in Luria–Bertani medium supplemented with ampicillin (100 µg.mL^{-}

[1]), under vigorous agitation (250 rpm) until Abs600 nm reached 0.6. Protein expression was induced by adding 1 mM isopropyl-β-D-thiogalactoside (IPTG) and cells were allowed to grow for 16 h at 18°C under vigorous shaking. Protein purification was performed with Co^{2+} columns according to the manufacturer's instructions (Talon Resin metal ion, Clontech). The buffer used to achieve the lysis of the cells contained 50 mM Hepes pH 7.6, 300 mM NaCl, 10 mM imidazole and 1 tablet of protease inhibitor cocktail (Sigma). The elution step was performed with 150 µM imidazole. Purified recombinant proteins were desalted against Hepes buffer (50 mM pH 7.6) using a PD-10 column (GE-Healthcare), concentrated with an Amicon®Ultra-4 (30 kDa cut-off, Merck) and conserved in 10% glycerol at -80°C upon usage. Protein concentration was determined according to the method described by [51] and purity was confirmed by separation on 12% acrylamide gels as described by [52]. Irrespective to the tested expression conditions, CRO_T104094 could not be produced as a soluble protein.

For the homodimerization analysis of the different isoforms, soluble crude protein extracts from *E. coli* expressing either 16OMT1 or 16OMT2 (CRO_T104105) were resolved in 10% SDS-PAGE after heating at 90°C for 5 min. Equal amount of non-heated protein extracts were resolved in 10% SDS-PAGE under semi-native conditions (loading buffer: 4% glycerol *v/v*, 240 mM Tris-HCL pH 6,8, 0.4% *w/v* blue bromophenol). The resolved proteins were then subjected to western blotting as described in [53] and [54].

2.9. Activity assays and kinetics experiments

Enzymatic assays were conducted in a total volume of 50 µL, by incubating 50 µM of the different substrates (tabersonine, 16-hydroxytabersonine, 19-hydroxytabersonine, vindoline, kaempferol, quercetin, naringenin), 1 mM S-(5')-adenosyl-L-methionine (SAM), 14 mM of β-mercaptoethanol and 2 µg of 16OMT1, CRO_T104094, CRO_T104105 or FlavoOMT recombinant protein in Hepes buffer (50 mM, pH 7.6) at 30°C during 60 min under slow agitation. Reactions were stopped by adding 1 *v/v* of 100% methanol, followed by vigorous shaking and centrifugation for 15 min at 16,000 ***g***, the supernatant was collected and subjected to UPLC-MS analysis. The reactions with flavonoids were stopped with 500 µL ethyl acetate, followed by the collection of the organic phase. The organic phase was evaporated to dryness using a Speedvac (ThermoFischer) and resuspended in 50% (*v/v*) methanol before analysis by UPLC-MS. A fresh protein fraction of cells harboring the empty vector of pRSET-A recovered after sonication was used as a negative control. For all reactions, three biological replicates were performed.

To determine kinetic parameters, the same reaction conditions were used as above,ccording to [55]. The concentration of the substrate 16-hydroxytabersonine varied between 0 and 20 µM, while the concentration of the cofactor SAM was set at 250 µM. All reactions were performed in three biological replicates. Reaction products (16-methoxytabersonine) were quantified by UPLC-MS according to a calibration curve. The kinetic parameters were calculated by non-linear regressions of the Michaelis–Menten plots using the software SigmaPlot12.

2.10. Raw protein extract activity assays

C. roseus tissues were grinded to fine powder in the presence of liquid nitrogen and crude proteins were extracted in ice-cold extraction buffer (50 mM HEPES-KOH, pH 7.5, 250 mM sucrose, 5% glycerol *v/v*, 10 mM EDTA pH 8, 0.5% PVP-10 v/v, 3 mM DTT, 1 mM PMSF) and recovered by centrifugation at 10,000 g for 10 min at 4 °C. The supernatant fraction was desalted against a 100 mM potassium phosphate buffer pH 7 containing 250 mM mannitol and 1 mM DTT, using a PD-10 column (GE Healthcare). The enzymatic assays were performed in a 200 µL total volume, by incubating 25 µM of 16-hydroxytabersonine, 0.5 mM SAM and 8 µg of total protein in 100 mM potassium phosphate buffer pH 7.6 at 30°C for 4 h. The reactions were stopped by adding 1 *v/v* of methanol 100 % and the results were analyzed by UPLC-MS. At least three biological replicates were performed.

2.11. UPLC-MS analyses

UPLC-MS analyses were conducted using equipment and conditions described in [2,14,50]. Data collection was carried in selected ion monitoring (SIM) mode for the following compounds: tabersonine m/z 337 RT = 14.4 min, 16-hydroxytabersonine m/z 353 RT = 8.4 min, 16-methoxytabersonine m/z 367 RT = 12.03 min, 19-hydroxytabersonine m/z 353 RT = 6.5 min, vindoline m/z 457 RT = 14.78 min, naringenin m/z 271 RT = 13.4 min, kaempeferol m/z 285 RT = 14.25 min, quercetin m/z 301 RT = 12.63 min, methylquercetin m/z 315 RT = 16.36 min, myricetin m/z 317 RT = 8.73 min, monomethylmyricetin m/z 331 RT = 11.8 min, syringetin m/z 345 RT = 15 min.

2.12. Subcellular localization studies and analysis of protein-protein interactions by bimolecular fluorescence complementation (BiFC)

The coding sequences of CRO_T104105 and CRO_T104094 were amplified by Phusion PCR using dedicated primers (**Supplementary Table S1**), which were designed to introduce *Spe*I

restriction sites to both extremities of the amplicon. The resulting product was cloned into either *Spe*I or *Nhe*I compatible restriction sites of the pSCA-vector YFPi [56] resulting in the fusion proteins CRO_T104105-YFP, YFP-CRO_T104105, CRO_T104094-YFP and YFP-CRO_T104094. The bimolecular fluorescence complementation (BiFC) assays were performed by cloning 16OMT1, CRO_T104105 and CRO_T104094 cDNAs into pSPYCE (MR) and pSPYNE173 vectors at the *Spe*I sites of either 5' or 3' ends of the N-terminal (YFP^N, residues 1-173) and C-terminal (YFP^C, residues 156-239) YFP fragments [57]. This allowed the expression of the fusion proteins 16OMT1-YFP^N, 16OMT1-YFP^C, CRO_T104105-YFP^N, CRO_T104105-YFP^C, CRO_T104094-YFP^N and CRO_T104094-YFP^C which were used to study protein interactions. The vectors constructed were transiently co-transformed with a plasmid harboring the nucleocytoplasmic CFP marker [34] into *C. roseus* cells by particle bombardment followed by fluorescence imaging [58]. The cells were bombarded with DNA-coated gold particles (1 μm) and a 1100-p.s.i. rupture disc at a stopping-screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He delivery system (Bio-Rad, Hercules, CA, USA). The cells were allowed to recover for 16-30 h before the harvesting and observation. The subcellular localization imaging was performed with an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera and a combination of YFP and CFP filters. The localization patterns of *circa* 50 observed cells were presented in this study.

2.13. Virus induced gene silencing (VIGS) in C. roseus plants

The pTRV1 and pTRV2 vectors used for generating VIGS constructs were procured from The Arabidopsis Information Resource (TAIR), USA. For VIGS of *16OMT1* and *16OMT2* (CRO_T104105), fragments of 520 bp and 531 bp were selected, respectively. The fragments were PCR amplified with gene-specific primers using *C. roseus* leaf cDNA (**Supplementary Table S1**). The resulting PCR products were purified and cloned into pTRV2 vectors. The resulting pTRV2::*Cr16OMT1* and pTRV2::*Cr16OMT2* plasmids were confirmed by sequencing and transformed into *Agrobacterium tumefaciens s*train GV3101 by the freeze-thaw method. A pTRV2 vector harboring a sequence common to both isoforms was also produced resulting in the pTRV2::OMTs constructs by using the primers described in **Supplemental Table S1**. The first silencing approach targeting each isoform separately, was performed according to [59]. For primary culture, a small scoop of frozen glycerol stock of *A. tumefaciens* harboring pTRV2, pTRV2::*CrOMT1* and pTRV2::*CrOMT2*, pTRV1, and pTRV2::*CrPDS* were inoculated into 5 mL of YEP broth supplemented with rifampicin (60 mg/L), gentamycin

(50 mg/L) and kanamycin (50mg/L). The culture was grown overnight at 200 rpm in an incubator shaker maintained at 28°C. For secondary inoculum, 500 µl of the primary culture was added to 50 mL YEP broth containing antibiotics rifampicin (60 mg/L), gentamicin (50 mg/L), kanamycin (50mg/L), and 200 µM acetosyringone, and grown overnight in an incubator shaker maintained at 28°C and 200 rpm. The overnight grown *A. tumefaciens* (secondary culture) was harvested at 5000 rpm for 5 min in a refrigerated centrifuge maintained at 4°C. The supernatant was discarded, and the bacterial pellet was resuspended in the infiltration buffer containing 10 mM MES (pH 7), 10 mM $Mgcl_2$ and 200 µM acetosyringone. The OD_{600} of the infiltration culture was adjusted to 1.6 and was further incubated at 28°C for 3-4 h at 200 rpm. Prior infiltration, *A. tumefaciens* cultures containing pTRV1 and pTRV2::*CrOMT1/CrOMT2/CrPDS* were mixed in a 1:1 ratio. For agroinfiltration, 4-6 leaf staged *C. roseus* seedlings were infected by pricking below the apical meristem using a dissecting needle. Post infiltration, the plants were kept in dark for 48 h to increase the virulence activity and then moved to the growth chamber (22 °C, 75% humidity, 16-8h light cycle). Leaves were harvested 21 days post infiltration and stored in -80 °C for further use. Plants infected with pTRV2::*CrPDS* were used to check the efficacy of VIGS in all experiments. For the common region silencing of both isoforms using the pTRV2::OMTs vector, a biolistic approach was used according to [46,60]. Plants with a first pair of fully developed leaves were selected and transformed with the PDS1000/He system (BioRAD). The bombardment was performed after the coating of the gold particles with the pTRV2::OMTs (1 µg) together with pTRV1 (1 µg) vectors. The empty vector pTRV2 was used as control. A single shot was made per plant. The plants were kept in a wet atmosphere for 24 hours post-bombardment, when a progressive re-acclimation was performed. At least, 5 transformed plants per condition were analyzed.

2.14. Transient overexpression in C. roseus plants

The pEAQ-HT vector was double digested with *Age*I/*Xma*I, followed by the respective cloning of the coding sequence of each candidate isoforms (FlavoOMT; 16OMT1; 16OMT2; CRO_T104094). The resulting plasmids as well as an empty vector (EV) were inserted into the *A. tumefaciens* cells by electroporation in the GV3101 (pMP90). Before infiltrations, Agrobacteria were prepared according to [39]. The plantlets were then infiltrated at a young stage using the vacuum technique and after 6 days post infiltration (6 dpi), the fluorescence was checked, and the fluorescent leaves were collected. The leaves were subjected to metabolic

and western blot analyses. The extraction of proteins and metabolites from transformed leaves as well the western blot conditions, were performed as described in [39]. The metabolite contents of the transformed plants were analyzed by UPLC-MS.

3. Results

3.1. Identification of 16OMT isoform candidates in C. roseus

To initiate the identification of potential new 16OMT candidates, we examined the latest annotated *C. roseus* genome [45] and retrieved 21 full-length coding sequences that were annotated as *O*-methyltransferases. A phylogenetic analysis showed that these sequences clustered into specific subclades including resveratrol-*O*-methyltransferases (CRO_T104083, CRO_T113075, CRO_T113071, CRO_T113068, CRO_T114313) and flavonoid-*O*-methyltransferases (CRO_T104109, CRO_T102208, CRO_T103893, CRO_T103890, CRO_T103897) (**Figure 2**). This last subclade notably includes the already identified CrOMT2 (CRO_T103893) and CrOMT4 (CRO_T103890), and is found in close proximity to a possible MIA specific subclade containing 16OMT1 (CRO_T110596) and the already characterized CrOMT5 (CRO_T104091) and CrOMT7 (CRO_T104106). Interestingly, we also identified CRO_T104105 and CRO_T104094 as the closest homologs of 16OMT1 (88.4% and 65.5% of identity, respectively). While the dimerization and *O*-methyltransferase domains were highly conserved in these candidates as well as in flavonoid *O*-methyltransferases, we noted that the residues predicted to be involved in the interaction with 16-hydroxytabersonine according to [61] slightly changed notably for Met 114, Trp 253 and Phe 312 (**Supplemental Figure 2**). Based on these results, we thus choose CRO_T104105 and CRO_T104094 for further characterization.

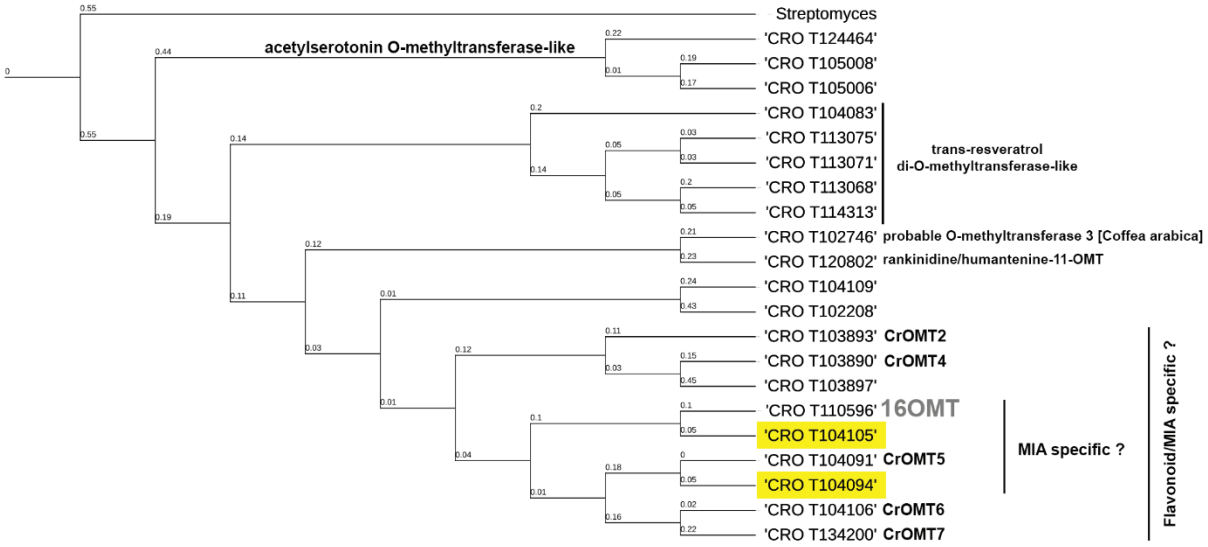


Figure 2. Neighbour-Joining tree of selected *C. roseus* OMT proteins. Streptomyces (methyltransferase domain-containing protein, NCBI: WP_191870180.1). Type II OMTs include *trans*-resveratrol di-*O*-mehtyltransferase-like (CRO_T104083), *trans*-resveratrol di-O-mehtyltransferase-like (CRO_T113075), *trans*-resveratrol di-O-mehtyltransferase-like (CRO_T113071), Probable *O*-methyltransferase 3 (CRO_T113068), *trans*-resveratrol di-*O*-mehtyltransferase-like (CRO_T114313), probable *O*-methyltransferase 3 [*Coffee arabica*] (CRO_T102746), rankinidine/humantenine-11-OMT (CRO_T120802), 67% id CrOMT7 (CRO_T104109), 57%id CrOMT7 (CRO_T102208), CrOMT2 (CRO_T103893), CrOMT4 CRO_T103890, CRO_T103897, 16OMT (CRO_T110596), CRO_T104105, CrOMT5 (CRO_T104091), CRO_T104094, CrOMT6 (CRO_T104106), CrOMT7 (CRO_T134200). Type I OMTs include Acetylserotonin O-methyltransferase-like CRO_T124464, CRO_T105008, CRO_T105006. Selected candidates are highlighted in yellow. Numbers in the internal branches denote the raw branch length values.

3.2. Functional validation of the 16OMT candidate isoforms

To assay for 16-*O*-methyltransferase activity, 16OMT, CRO_T104105 and CRO_T104094 were expressed in yeast by cloning their coding sequences into autoreplicative plasmids bearing inducible promoters (pESC). While CRO_T104104 could be easily amplified from young leaf RNA, all our attempts to clone CRO_T104094 from retrotranscribed RNA failed. Given the limited intron number located within is genomic sequence, the entire CRO_T104094 was amplified from genomic DNA, cloned into the pEAQ-HT binary vector for agroinfiltration mediated expression in *N. benthamiana* leaves (**Supplemental Figure 1**). Such procedure allowed the amplification of an intron-less sequence from *N. benthamiana* leaf RNA and its cloning into the yeast expression vector. The three resulting plasmids (pESC-16OMT; pESC-CRO_T104105; pESC-CRO_T104094) were individually transformed with a distinct plasmid

expressing T16H2 into yeast. Two additional strains transformed by a pESC empty vector and pYePD60-T16H2 alone were also generated. The transformed yeast strains were cultured and induced with galactose before feeding with tabersonine (100 µM) for 0.5, 3 or 16 h and analysis of the resulting products was performed by UPLC-MS.

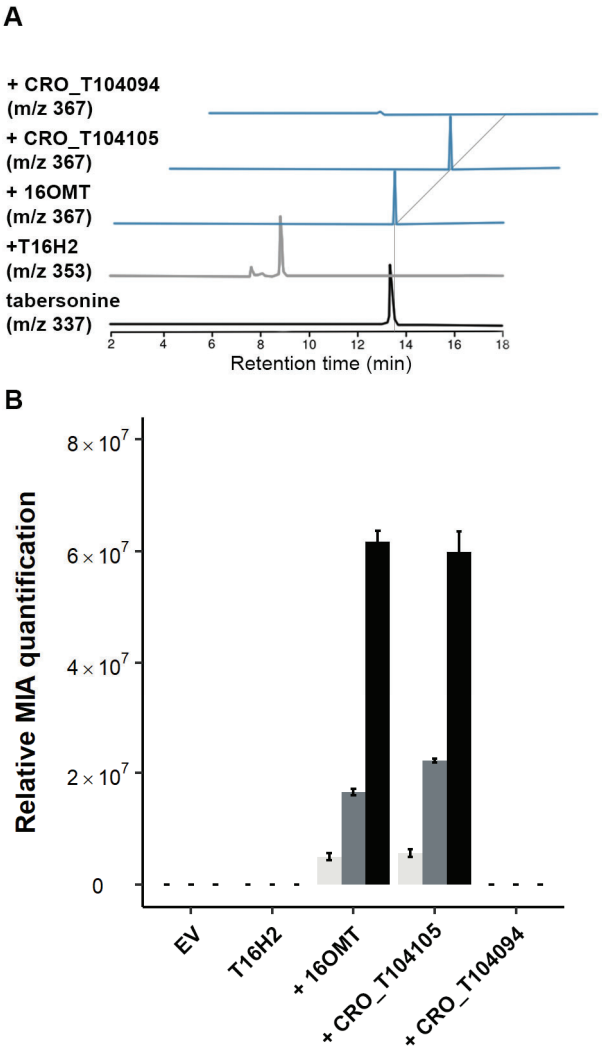


Figure 3. Screening selected candidates for tabersonine 16-*O*-methyltransferase activity. (A) MIA biosynthetic intermediates in the medium culture of galactose induced yeast harboring episomal plasmids. Yeast transformed by an empty vector (EV) or expressing T16H2 (grey line) with selected 16OMT-like candidates (CRO_T104105 and CRO_T104094) and positive control 16OMT1 (blue lines) were incubated with tabersonine (black line, m/z 337) for 16 hours. Selected ion monitoring and comparisons to authentic standards were used to detect the resulting hydroxylated (m/z 353) and methylated (m/z 367) tabersonine products by UPLC–MS analysis. The light grey line highlights 16-methoxytabersonine retention time. (B) Evolution of the 16-methoxytabersonine content in each yeast strain over time. Tabersonine and 16-methoxytabersonine were quantified by UPLC-MS in the yeast culture medium 0.5 (light grey bars), 3 (dark grey bars) and 16 h (black bars) post-feeding with tabersonine (250 µM). Error bars correspond to the standard deviation of biological replicates (n = 3). MIA composition of the yeast culture medium is expressed as relative peak areas.

While yeasts transformed by empty vector and pYeDP60-T16H2 did not metabolize tabersonine and produce 16-hydroxytabersonine respectively, yeasts expressing both T16H2 and 16OMT synthesized a compound with a mass-to-charge ratio (*m/z*) 367 corresponding to 16-methoxytabersonine (**Figure 3A**). Interestingly, yeasts transformed by pYeDP60-T16H2 and pESC-CRO_T104105 produced a compound with a similar m/z, retention time and fragmentation pattern (**Supplemental Figure S3**) that likely also corresponds to 16-methoxytabersonine. We also observed that the amount of reaction products generated by the two strains remained in the same order of magnitude along time (**Figure 3B**). Conversely, no 16-methoxytabersonine was generated by yeast co-expressing T16H2 and CRO_T104094 and no consumption of 16-hydroxytabersonine was observed strongly suggesting that CRO_T104094 did not display any activity towards this substrate (**Figure 3A-B**). A closer look at the CRO_T104094 sequence revealed a gap between the amino acid 84 and the amino acid 87 compared to 16OMT and CRO_T104105 (**Supplemental Figure S4A**). This deletion is possibly related to a twist between the □-sheet 2 and 3 resulting in a shortening of the □-sheet 2, according to protein modelization (**Supplemental Figure S4B**). To determine whether this could explain the lack of activity of CRO_T104094, a chimeric version of the protein was generated by PCR in order to integrate the missing residues from 16OMT. However, after the cloning of the corresponding sequence in pESC and expression in yeast with T16H2, we did not detect any 16-hydroxytabersonine methyltransferase activity, thus establishing that the missing residues do not explain the lack of enzyme activity on 16-hydroxytabersonine (**Supplemental Figure S4C-D**). Definitively, all these results strongly suggested that CRO_T104105 potentially corresponds to a second isoform of 16OMT. We thus renamed 16OMT as 16OMT1 and CRO_T104105 as 16OMT2 hereafter.

3.3 16OMT1 and 16OMT2 display similar kinetic parameters but distinct substrate preferences

To gain insight into the enzymatic specificities of 16OMT1 and 16OMT2, we first analyzed their biochemical properties towards different substrates after expressing and purifying the corresponding recombinant enzymes in *E. coli*. Given the phylogenetic relationship between 16OMTs and flavonoid-*O*-methyltransferase (**Figure 2**), we also compared both proteins with a previously characterized flavonoid 4-*O*-methyltransferase (FlavoOMT, AY127568, OMT4, [33]). For substrate specificity assays, each recombinant enzyme was incubated with distinct MIAs (tabersonine, 16-hydroxytabersonine, 19-hydroxytabersonine and vindoline) or

flavanones and flavonols (naringenin, kaempferol, quercetin, myricetin and monomethylmyricetin) in the presence of SAM (**Table 1**). As expected, 16OMT1 and 16OMT2 methylated with very high efficiency 16-hydroxytabersonine but did not accept tabersonine as substrate, given the lack of the hydroxyl group (**Table 1, Figure 4**). Both enzymes neither methylated 19-hydroxytabersonine nor vindoline confirming the specificity of methylation at the C16 position. Interestingly, we saw that FlavoOMT used 16-hydroxytabersonine with a very low efficiency, around 2.5% of substrate converted, but no other tested MIAs (**Table 1, Figure 4**). By contrast, the FlavoOMT displayed high catalytic activities against flavonols and notably quercetin, myricetin and monomethylmyricetin, as previously reported [33]. Interestingly, while 16OMT1 exhibited barely detectable activities toward quercetin, myricetin, 16OMT2 substantially methylated myricetin (8% conversion rate) and quercetin (25% conversion rate) thus suggesting that it displays a broader substrate specificity than 16OMT1 and FlavoOMT.

Table 1. *16OMT1 and 16OMT2* substrate specificities. FlavoOMT was assayed as the reaction control.

+ and – denote capacity and incapacity of methylation. The conversion of substrates (%) is given in brackets.

Substrate	**16OMT1**	**16OMT2**	**FlavoOMT**
Naringenin	-	-	-
Kaempferol	-	-	-
Quercetin	+ (0.84 ± 0.19)	++ (25.31 ± 0.03)	++ (45.96 ± 0.01)
Myricetin	+ (1.19 ± 0.30)	+ (8.32 ± 0.02)	+++ (70.15 ± 0.08)
Monomethylmyricetin	-	-	+++ (74.88 ± 0.04)
Tabersonine	-	-	-
16-Hydroxytabersonine	+++ (100 ± 0.02)	+++ (99.94 ± 0.04)	+ (2.47 ± 0.15)
19-Hydroxytabersonine	-	-	-
Vindoline	-	-	-

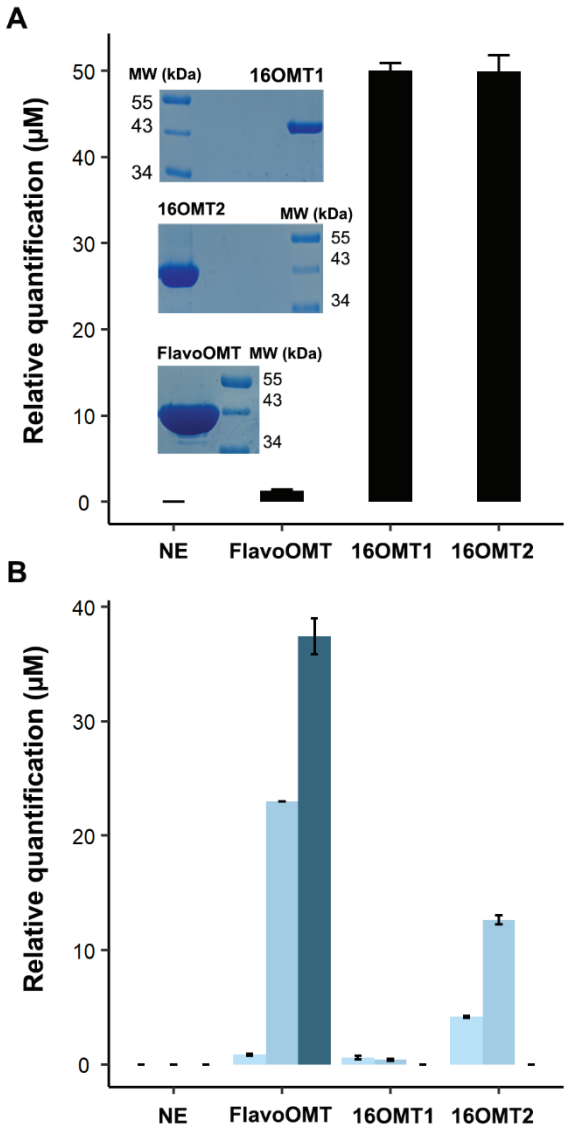


Figure 4. Relative activity of 16OMT1, 16OMT2 and FlavoOMT against 16-hydroxytabersonine, quercetin and myricetin. Activities of the recombinant FlavoOMT, 16OMT1 and 16OMT2 were determined by incubation with 16-hydroxytabersonine (A) or quercetin and myricetin (B) and SAM for 2 h and compared with the same reaction performed with no enzyme (NE). The purity of the recombinant 16OMT1, 16OMT2 and FlavoOMT was confirmed on acrylamide gel stained with Coomassie blue. For 16-hydroxytabersonine, the formation of 16-methoxytabersonine (dark) was monitored but no differences in conversion were observed between 16OMT1 and 16OMT2 after a 2 h assay. For flavonoids, a methylated product was observed in the presence of quercetin (light blue) and myricetin (medium blue) for the three proteins. The FlavoOMT was able to methylate myricetin a second time to generete syringetin (dark blue). Error bars correspond to the standard deviation of biological replicates (n = 3). Products composition are expressed as relative peak areas in micromolar.

Finally, we evaluated the kinetic parameters of both 16OMT1 and 16OMT2 towards their preferred substrate namely 16-hydroxytabersonine. Each recombinant enzyme was thus incubated with increasing amounts of 16-hydroxytabersonine as a methyl acceptor and SAM as a methyl donor. After plotting the results, the Michaelis-Menten model was fitted to the points by non-linear regression (**Figure 5**). Under these conditions, substrate saturation kinetics revealed that 16OMT1 and 16OMT2 share similar catalytic properties with a high affinity for 16-hydroxytabersonine (16OMT1, Km 0.52 ± ± $4.31.10^{-2}$ µM and Vmax 0.0014 ± $3.05.10^{-5}$ $\mu M.s^{-1}$; 16OMT2, Km 1.7 ± 0.15 µM and Vmax 0.0081 ± $2.42.10^{-4}$ $\mu M.s^{-1}$). All these results thus confirm that 16OMT2 truly corresponds to a new isoform of MIA biosynthetic enzyme displaying similar enzymatic specificities than 16OMT towards 16-hydroxytabersonine but a broader substrate specificity notably accepting flavonols.

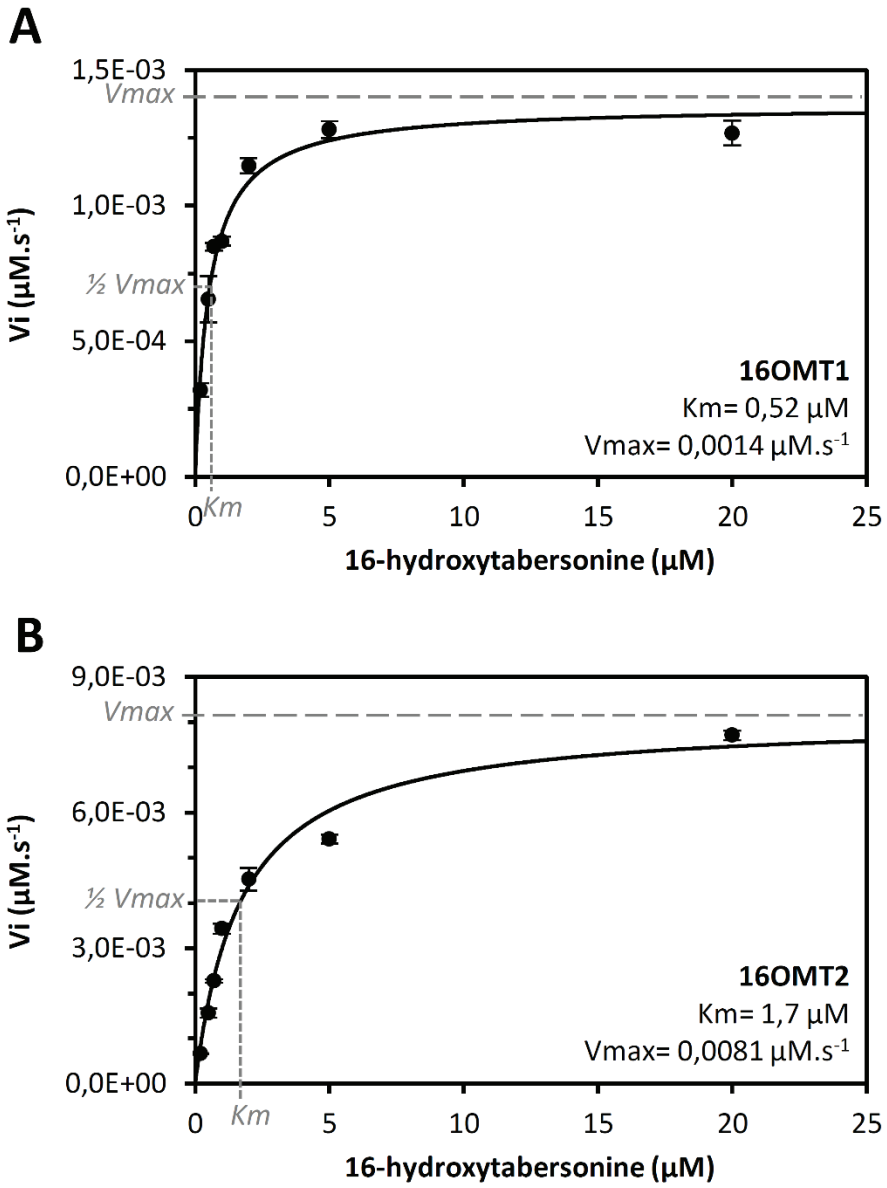


Figure 5. Determination of 16OMT1 (A) and 16OMT2 (B) kinetic parameters for 16-hydroxytabersonine. 16OMT1 and 16OMT2 initial velocities in 16-methoxytabersonine synthesis were measured using a fixed SAM concentration (250 µM) and 0 to 20 µM 16-hydroxytabersonine. Nonlinear regressions based on the Michaelis–Menten equation were applied to estimate kinetic parameters (16OMT1, R^2= 0,964 ; 16OMT2, R^2=0,976). Error bars correspond to standard deviation of biological replicates (n = 3).

3.4 16OMT isoforms operate as homodimers in the cytosol.

While OMTs mostly form homodimers, the recent discovery of the formation of OMT heterodimers impacting the synthesis of BIA prompted us to study 16OMT1 and 16OMT2 interactions. As a first step, we characterized the subcellular localization of 16OMT2 through YFP imaging. The coding sequence of 16OMT2 was cloned at the 5'- or 3'-end of the YFP coding sequence into the pSCA-YFP plasmid to express the 16OMT2-YFP and YFP-16OM2 fusion proteins, respectively. These two constructs were then transiently expressed in *C. roseus* cells in combination with a CFP nucleocytosolic marker (**Figure 6A-H**). Interestingly, both proteins displayed a similar signal of fluorescence that partially merged with that of the CFP marker. While a perfect superimposition of signals emanating from the cytosol was observed, no overlaps of signals from the nucleus were monitored (**Figure 6C, G**). The false red signal generated by the CFP marker in co-transformed cells thus suggested that 16OMT2 was not located in this subcellular compartment but rather mostly remained in the cytosol.

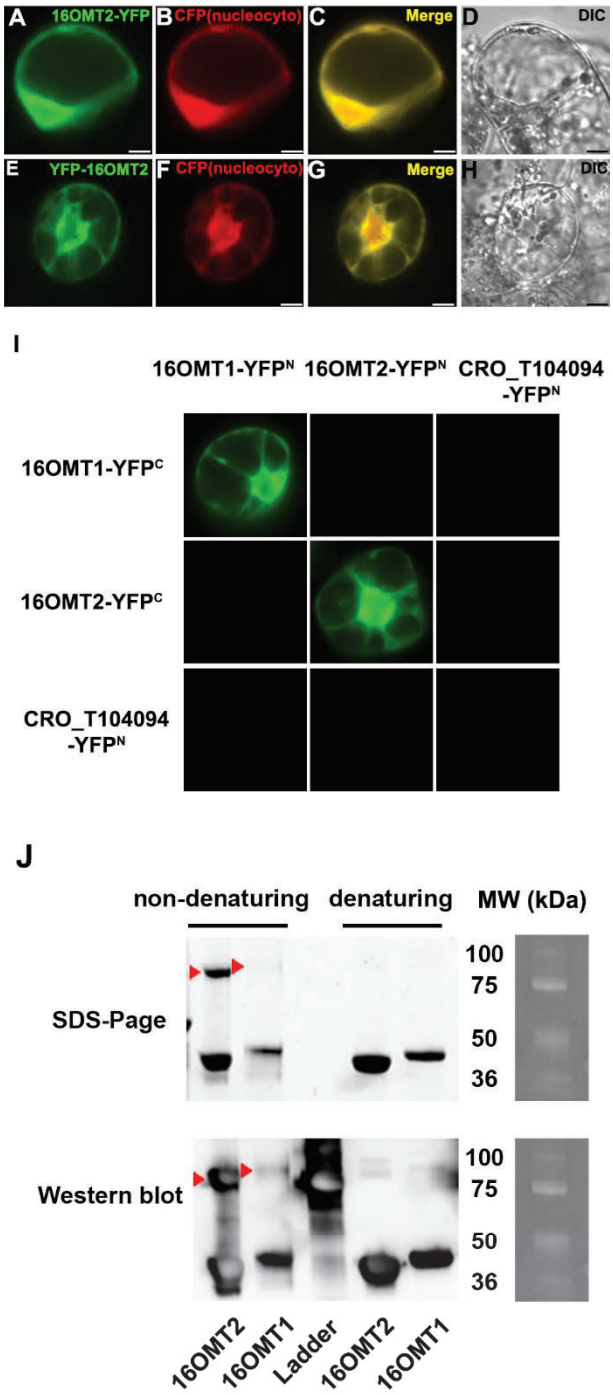


Figure 6. 16OMT1 and 16OMT2 forms homodimers in the cytosol of *C. roseus* cells. (A) Subcellular localization of 16OMT2 in *C. roseus* cells. Cells were transformed transiently with 16OMT2-YFP (A–D), YFP-16OMT2 (E-H) expressing vectors in combination with a CFP-nucleocytosolic marker as mentioned on the images of the second column. Co-localization of the two fluorescence signals appeared on the merged image (C, G). The morphology is observed with differential interference contrast (DIC). Bar: 10 µm. (I) Analysis of 16-OMT1, 16OMT2 and CRO_T104094 by BiFC assays. *C. roseus* cells were transiently co-transformed using a combination of plasmids as indicated on the top (fusion with the YFP^N fragment) and on the left (fusion with the YFP^C fragment).

Figure 6 (continuation). (J) SDS-Page and western blot analyses of self-interactions between recombinant 16OMT1 and 16OMT2. The recombinant purified proteins were analyzed under denaturing conditions (SDS-PAGE) or under non-denaturing conditions, to show the presence and lack of homodimers. A Westen-blot performed on the same samples with an anti-6His antibody was also performed. Red arrowheads highlight protein homodimers.

On this basis, we next determined whether interactions could occur between both 16OMT isoforms through the analysis of their quaternary structures by bimolecular fluorescence complementation (BiFC) assays. Based on the head-to-head model of interaction previously characterized for 16OMT1 [34], the 16OMT2 ORF was cloned at the 5'-end of the coding sequence of the split-YFPN and split-YFP^C fragments to generate 16OMT2-YFP^N and 16OMT2-YFP^C proteins, respectively. The similar constructs already generated for 16OMT1 were also used [34]. As a negative control, we cloned CRO_T104094 in the same plasmids to generate the CRO_T104094-YFP^N and CRO_T104094-YFP^C constructs, after having established the cytosolic localization of this enzyme (**Supplemental Figure S5**). After transient expression in *C. roseus*, we firstly established that 16OMT2 formed homodimers, at least, based on the acquisition of a YFP fluorescent signal following YFP reformation due to 16OMT2 self-interactions (**Figure 6I**). As expected, a similar cytosolic signal was observed for 16OMT1 in agreement with our previous observations [34]. Conversely, no signal was acquired for CRO_T104094 thus suggesting a lack of homodimer formation which could be correlated with its lack of activity (**Figure 6I**). This formation of 16OMT2 dimers was then confirmed by analyzing the interaction of the recombinant enzyme through electrophoretic migration in PAGE under denaturing and non-denaturing conditions (**Figure 6J**). While migration under denaturing conditions revealed a unique protein band for both 16OMT1 and 16OMT2, non-denaturing conditions allowed the observation of a band of a higher size for both enzymes that likely results from protein homodimerization. Such observation was confirmed through Western blot analysis by taking advantage of the presence of a 6xHis-tag at the N-terminal end of 16OMT1 and 16OMT2. Finally, BiFC assays were conducted by mixing the 16OMT1,16OMT2 and CRO_T104094 different YFP split variants. However, in any of the tested associations we observed the formation of YFP fluorescent signal thus strongly suggesting that 16OMT isoforms do not form heterodimers. No signal was obtained from associations including CRO_T104094. These results thus suggest that both 16OMT1 and 16OMT2 specifically formed homodimers in the cytosol.

3.5. 16OMT1 and 16OMT2 display different gene expression profiles among C. roseus organs

To gain insights into the possible involvement of 16OMT2 in MIA biosynthesis, we analyzed its gene expression and compared it to that of 16OMT1 and with 16OMT activity and 16-methoxytabersonine accumulation *in planta*. Both transcript levels were quantified in the major *C. roseus* organs and revealed distinct expression patterns (**Figure 7A**). As previously observed [34], 16OMT1 expression reached a maximum in young leaves, remaining high in almost all the tested aerial organs except stems but was not detectable in roots. By contrast, 16OMT2 transcripts were detected in all organs including roots, reaching very high amounts in fruits, flowers, fully developed leaves and stems notably and lowest levels in roots and young leaves. Overall, 16OMT2 expression was around 10-fold higher than that of 16OMT1 both genes expressed a similar transcription level in young leaves. In addition, we also noted that 16OMT2 expression was specifically located in the epidermis of leaves in agreement with the MIA pathway architecture as previously observed for 16OMT1 (**Supplemental Figure S6**, [9,34]). Interestingly, 16OMT activity did not strictly follow these gene expression profiles but was detected in all organs (**Figure 7B**). While 16OMT activity in roots and stems could potentially be correlated to the sole 16OMT2 expression, it rather resulted from both gene expressions in all other organs with a possible main involvement of 16OMT2 given its highest expression. We also noted that 16OMT activity reached a maximum in young leaves and flower buds. While this was in agreement with the high vindoline metabolism in these organs, it differed from the cumulated amount of 16OMT1 and 16OMT2 transcripts suggesting that other mechanisms/isoforms could tightly regulate 16-hydroxytabersonine methoxylation. Such a reasoning was supported by 16-methoxytabersonine quantification (**Figure 7C**), which revealed a low accumulation of this MIA intermediate except in young leaves to fulfill vindoline synthesis. All these results thus established that 16OMT2 displays an expression pattern that partially follows 16OMT activity and MIA synthesis.

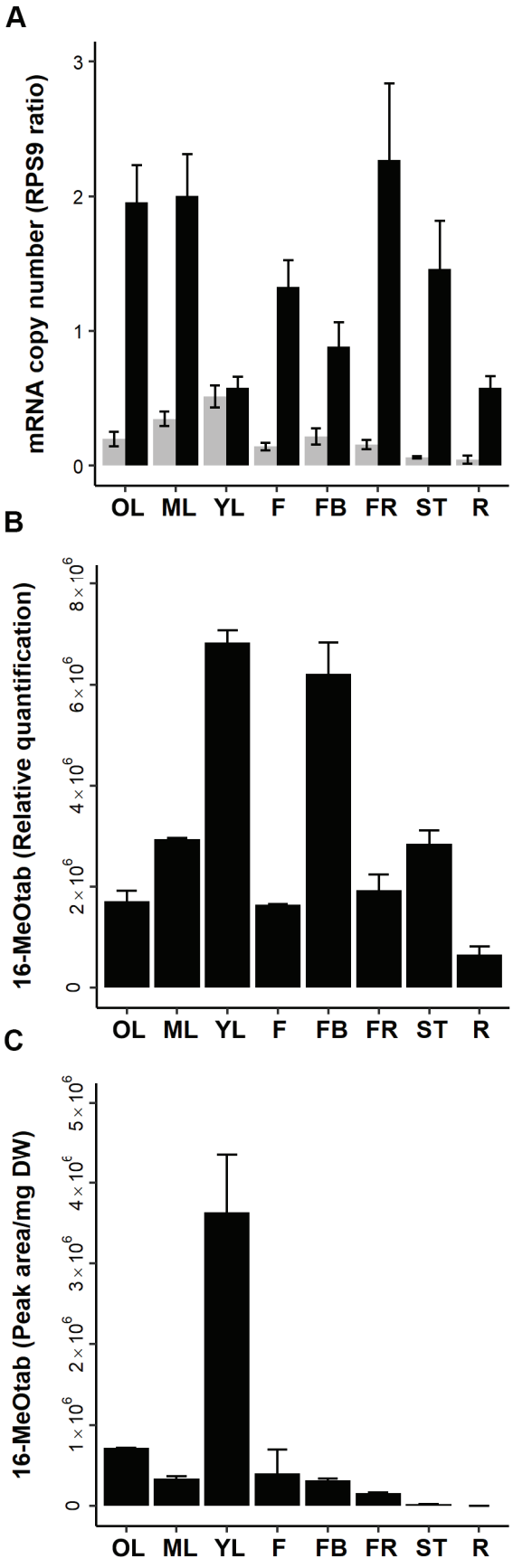

A
mRNA copy number (RPS9 ratio)
0
1
2
3
OL ML YL F FB FR ST R
B
16-MeOtab (Relative quantification)
0
2×10^6
4×10^6
6×10^6
8×10^6
OL ML YL F FB FR ST R
C
16-MeOtab (Peak area/mg DW)
0
1×10^6
2×10^6
3×10^6
4×10^6
5×10^6
OL ML YL F FB FR ST R

Figure 7. Comparison of 16OMT1/16OMT2 expression with 16OMT activity and accumulation of 16-methoxytabersonine in different *C. roseus* organs. (A) 16OMT1 and 16OMT2 transcript quantifications were performed by reverse transcription quantitative PCR (qPCR) using RNA extracted from different organs, and expressed relatively to RPS9 used as a control gene. Grey bars: 16OMT1 and black bars: 16OMT2. (B) Raw protein extracts (8 µg) from each *C. roseus* organ were co-incubated with 16-hydroxytabersonine and SAM during 4 h at 30°C. The values of 16-methoxytabersonine (16-MeOtab) are expressed in relative units. (C) Quantification of 16-methoxytabersonine amounts (16-MeOtab) in each organ of *C. roseus* by UPLC-MS. Old leaves (OL), Medium leaves (ML), Young Leaves (YL), Flowers (F), Flower buds (FB), Fruits (FR), Stems (ST) and Roots (R). MIA composition is expressed as relative peak areas. Error bars correspond to standard deviation of biological replicates (n = 3).

3.6 Silencing and overexpression confirm the involvement of 16OMT2 in vindoline synthesis

To investigate the role of 16OMT2 in vindoline biosynthesis *in planta*, we performed genomic functional analyses by first applying a transient silencing approach through virus-induced gene silencing (VIGS). The high identity between 16OMT1 and 16OMT2 prompted us to try identifying a partial sequence allowing a specific silencing of each isoform. However, a first assay based on the amplification of two 500-bp sequences located at the 5'-end of their coding sequences followed by their cloning into the pTRV2 plasmid and transformation of young *C. roseus* plantlets resulted in cross-silencing reactions (**Supplemental Figure S7, S8**). This established that a specific downregulation of 16OMT1 and 16OMT2 was presumably not achievable by VIGS, leading us to perform a 16OMT1/16OMT2 co-silencing. A 426-bp sequence displaying 90 % of identity was thus selected to generate the 16OMT co-silencing construct (**Supplemental Figure 9**). Young plantlets were then transformed by this construct or an empty vector (EV) as a control. Around three weeks post-transformation, we measured a pronounced reduction of both 16OMT1 and 16OMT2 transcription levels in *C. roseus* leaves, which dropped down to around 80% of those measured in EV-transformed plants thus confirming both 16OMT isoform cross-silencing. By contrast, no effect was observed on a distinct MIA biosynthetic gene encoding strictosidine synthase (STR) confirming silencing specificity (**Figure 8B**). Such gene silencing was accompanied by a decrease of 16-methoxytabersonine synthesis (around 75 %) and lower vindoline accumulation (around 55 %) thus suggesting that 16OMT1 and 16OMT2 are both important drivers of MIA synthesis (**Figure 8B**).

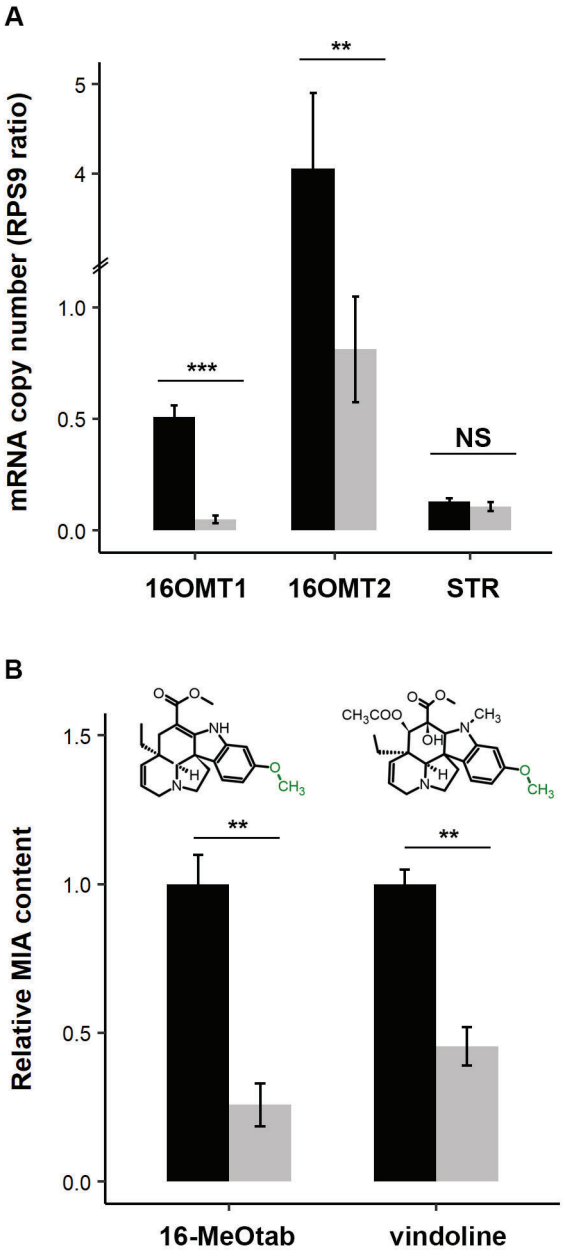


Figure 8. **Down-regulation of 16OMT affects vindoline biosynthesis.** (A) 16OMT isoforms transcript down-regulation by VIGS was determined by real-time RT-PCR analyses performed on total RNA extracted from *C. roseus* leaves of 16OMT1/2-silenced plants (gray bars) or empty-vector control transformed plants (black bars). CrRPS9 was used as a housekeeping gene and STR (strictosidine synthase) was used as a control for the pathway. Data correspond to average values (n = 5) ± se of independent transformed plants. (B) Relative MIA content in 16OMT1/2-silenced plants (gray bars) compared to MIA content in plants transformed with empty vector (black bars). The relative amounts of 16-methoxytabersonine (16-MeOtab) and vindoline were quantified by UPLC-MS. The amount of each MIA was expressed relative to that measured in empty vector plants (normalized to 1). Asterisks denote statistical significance (*P < 0.05, **P < 0.01, ***P < 0.001 by Student's t test). The results shown are representative of five OMTs-VIGS plants and five empty vector plants.

Finally, in order to get compelling evidence of 16OMT2 involvement in 16-hydroxytabersonine methoxylation in *C. roseus* leaves, we performed a complementary analysis through transient overexpression assays. The 16OMT2 coding sequence was thus cloned into the pEAQ-HT plasmid to produce the pEAQ-HT-16OMT2 expression vector. Likewise, the 16OMT1, CRO_T104094 and FlavoOMT coding sequences were cloned to generate the pEAQ-HT-16OMT1, pEAQ-HT- CRO_T104094 and pEAQ-HT-FlavoOMT expressing vectors. An empty vector was also used as a negative control. Following individual transformation of Agrobacterium, the resulting strains were used to perform Agroinfiltration of *C. roseus* leaves in combination with a pEAQ-HT-GFP vector expressing the green fluorescent protein. One week after infiltration, leaves expressing highly GFP were collected, and proteins and MIA extracted for analyses. Firstly, the overexpression of each of the tested protein was confirmed by Western blot analysis using an anti-His antibody. While the levels of expression vary depending on the construct, we confirmed the overexpression of each protein (**Supplemental Figure S10**). At the MIA level, we saw that the amounts of 16-methoxytabersonine and vindoline were similar in leaves overexpressing EV and CRO_T104094, thus confirming that this enzyme does not display any direct activity towards 16-hydroxytabersonine in agreement with our previous biochemical assays (**Figure 9A-B**). Similarly, despite the low activity measured in vitro, the overexpression of FlavoOMT did not result in any increase of 16-methoxytabersonine and vindoline content, strongly suggesting that FlavoOMT rather plays a role in flavonoid methylation *in planta*. By contrast, overexpressing 16OMT1 and 16OMT2 led to a strong and similar increase of both MIAs (of around 2-fold) confirming an active role of both enzymes in vindoline synthesis. Finally, we noted that these increases in 16-methoxytabersonine and vindoline were similar to those resulting from T16H2 overexpression, which could suggest a role of a same importance for both T16H2 and 16OMTs catalyzed reactions in the vindoline biosynthesis flux (**Supplemental Figure S11**).

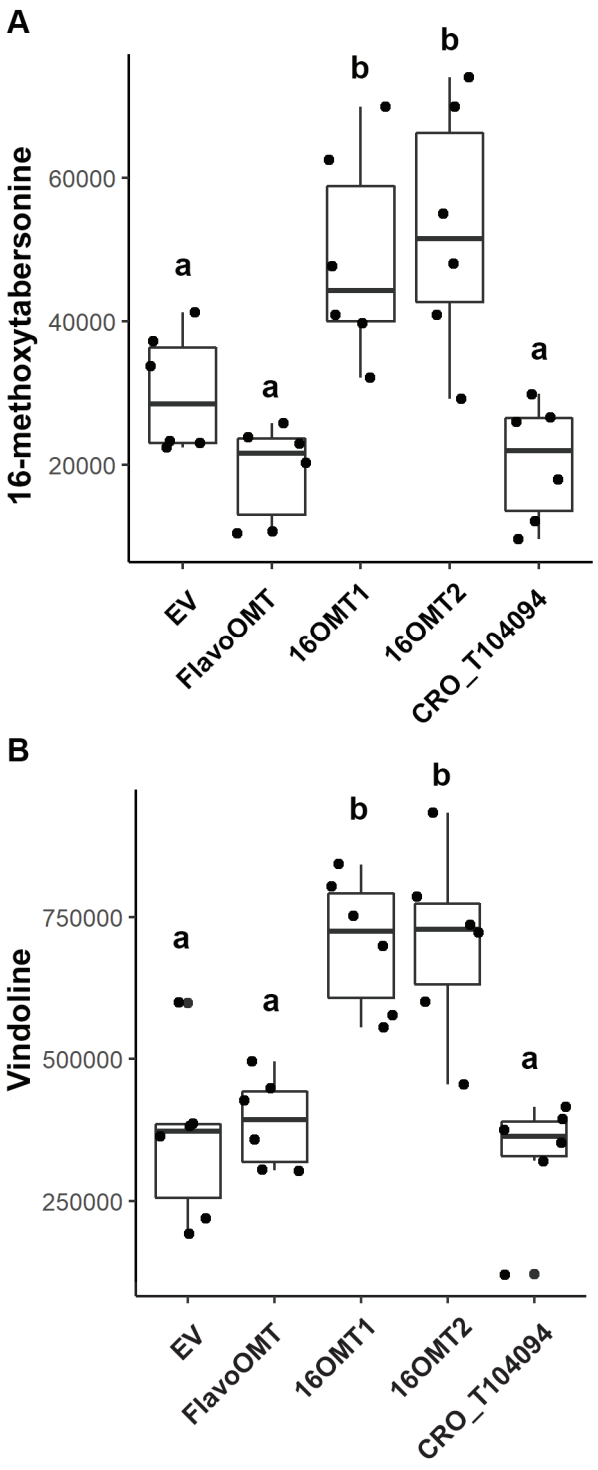


Figure 9. MIA accumulation in plants overexpressing empty vector (EV), FlavoOMT, 16OMT1, 16OMT2 and CRO_T104094. The 16-methoxytabersonine (A) and vindoline (B) concentration in *C. roseus* infiltrated leaves was measured 7 days after infiltration. Points correspond to individual samples, in which 2 leaves per plant were analyzed (3 plants were transformed for each condition). The individual samples were analyzed by ANOVA followed by the post hoc Tukey test. Same letters express no significant difference between the means of the conditions at 5% of significance. MIA composition is expressed in relative unit (Peak area/leaf dry weight).

4. Discussion

In *C. roseus*, the biosynthesis of MIAs results from long and complex pathways involving different leaf tissue-types and organelles but also several regulatory mechanisms [9,10]. Over the last years, the prominent role of enzyme isoforms in MIA biosynthetic fluxes has been highlighted notably for the synthesis of secologanin [10], tetrahydroalstonine [40,41] as well as for the first step of vindoline production catalyzed by T16H1 and T16H2 [42]. While recent reconstitutions of the tabersonine to vindoline pathway in yeast have revealed the prime importance of tabersonine 16-methoxylation for a high MIA production [15,16], only a minor attention has been paid to the reaction catalyzed by 16OMT1 after its cloning achieved more than 13 years ago [29].

By combining homology searches based on the previously characterized 16OMT1 and phylogenetic analyses, we have identified potential new 16OMT candidates encoded by CRO_T104105 and CRO_T104094, as well as the already characterized flavonoid-*O*-methyltransferases [32,33]. All these OMTs belong to the clade II OMTs and form a specific subclade mainly composed of flavonoid-*O*-methyltransferases. Functional assays in yeast combined to biochemical characterization revealed that CRO_T104105 encoded a *bona fide* 16OMT, and was named 16OMT2, while no activity towards the tested substrates was observed for CRO_T104094. Such a 16-hydroxytabersonine *O*-methyltransferase activity of 16OMT2 was consistent with its very high identity with 16OMT1 (around 88%) and strongly supported the identification of a second isoform of 16OMT as previously observed for the previous step of this biosynthetic pathway [42]. However, while 16OMT1 and FlavoOMT are highly specific toward their respective substrate, 16OMT2 appears to be a multifunctional enzyme working on both MIA and flavonoid substrates (Table 1). Such properties thus resemble those of the recently identified *Camptotheca accuminata* 10-hydroxycamptothecin *O*-methyltransferase (Ca10OMT) also accepting both types of substrates [62]. For Ca10OMT, this substrate promiscuity has been explained by the structural similarities of 10-hydroxycamptothecin and flavonoids, which have constrained evolution of the enzyme to accommodate the MIA substrate and retaining a high flavonoid-*O*-methyltransferase activity. While the structural similarity between 16-hydroxytabersonine and flavonoids is less pronounced, a similar mechanism of evolution has probably occurred to favor the apparition of 16OMT2 from an ancestral specific flavonoid-*O*-methyltransferase. This hypothesis is supported by the genomic neighboring of 16OMT2 enriched in flavonoid-*O*-methyltransferases, chalcone and stilbene

synthase as well as CRO_T104094, which could reflect local gene duplication followed by neofunctionalization (**Supplemental Figure S12**). Along evolution, the 16OMT2 ancestor has probably lost its activity towards flavonoid and conversely increased that for 16-hydroxytabersonine, in agreement with the biochemical properties of 16OMT2 and FlavoOMT for instance. Interestingly, most of the characterized alkaloids *O*-methyltransferases display a high specificity towards their respective substrate as observed for coronaridine OMT [63], ipecac OMT [64], benzylisoquinoline alkaloid OMTs [65,66] and 16OMT1 (Table 1;[29]). Since 16OMT1 and 16OMT2 displays a high sequence identity and similar kinetic parameters for 16-hydroxytabersonine, it is reasonable to consider that 16OMT1 may result from a later 16OMT2 gene duplication accompanied by a continuous neofunctionalization leading to an almost complete loss of flavonoid *O*-methyltransferase activity. Surprisingly, the 16OMT1 encoding gene is located just next to T16H1 and T16H2, thus forming one of the sole MIA gene cluster identified so far (Supplemental Figure S12, [35]). This organization probably favors a coordinated gene expression to tightly regulate tabersonine methoxylation.

In agreement with their similar 16-hydroxytabersonine methyltransferase activity, the functional characterization of 16OMT1 and 16OMT2 strongly supports their involvement in vindoline biosynthesis *in planta*. While the high sequence identity did not allow to perform a specific gene down-regulation, the co-silencing led to a high decrease of both 16-methytabersonine and vindoline accumulation suggesting that both 16OMT1 and 16OMT2 could be involved in the synthesis of these MIAs in leaves. While gene silencing in *C. roseus* never completely abolished MIA synthesis [14,46,67–69], the remaining synthesis of vindoline in silenced plants may also reflect the involvement of yet not discovered 16OMT isoforms. On the other hand, the individual overexpression of 16OMT1 and 16OMT2 provided compelling evidence of their role in MIA synthesis (**Figure 9**, **Supplemental Figure S11**) in agreement with their gene expression profiles. As previously observed for 16OMT1 [29,34], we saw that in leaf 16OMT2 was mainly expressed in epidermis in agreement with the production of 16-hydroxytabersonine in this specific tissue type (**Supplemental Figure S6**; [42,50]). In addition, the strong expression of 16OMT2 in all aerial organs could thus be associated with its dual MIA and flavonoid *O*-methyltransferase activity, given the accumulation in the corresponding compounds [70–72]. It may thus be inherited from the ancestral parent gene encoding a *bona fide* flavonoid *O*-methyltransferase metabolizing compounds accumulated in the whole plant including roots. By contrast, the expression of 16OMT1 appeared different with a lower expression and a maximum reached in young leaves (**Figure 7A**; [34]). This may also reflect

the progressive specialization of this gene in MIA synthesis allowing it to follow vindoline accumulation and T16H2 expression profile. Interestingly, while 16OMT2 is always expressed at a higher level, the amount of 16OMT1 and 16OMT2 transcripts is similar in young leaves. Beside their similar biochemical properties, this supports the idea that 16OMT1 and 16OMT2 could have a similar importance in the foliar synthesis of vindoline. By contrast, 16OMT2 could be of a primer importance in the floral vindoline production, since around 10-fold more expressed than 16OMT1. Such expression profiles thus suggest that both 16OMT isoforms could display both redundant and complementary roles in the synthesis of vindoline depending on organ-type.

Finally, we established that 16OMT2 forms homodimers in the cytosol as also described for 16OMT1 and 16OMT from *Vinca minor* (Figure 6; Figure 7; [34,55]). Such conformation allows to prevent its passive diffusion into the nucleus to favor the uptake of the 16-hydroxytabersonine released in the cytosol by the endoplasmic reticulum-anchored T16H2 [42]. However, we did not observe any formation of heterodimers between 16OMT1 and 16OMT2 under our tested conditions. While we cannot exclude the possibility that other functional undiscovered 16OMT could be involved in such a process, this situation seems to differ from that described for the BIA synthesis where OMT cross interactions allow the diversification of the BIA production [43]. In *C. roseus*, the sole formation of 16OMT1 or 16OMT2 homodimers probably results from a high specify of interactions but also potentially reflects a lower requirement of diversification of MIA methylation. Indeed, several *O*-methylation steps are required for BIA synthesis while only one is involved in the tabersonine to vindoline conversion. Finally, such a lack of interaction also suggests that 16OMT1 and 16OMT2 play complementary roles in vindoline synthesis.

Declaration of Competing Interest

The authors have declared that no competing interests exist.

Acknowledgements

We acknowledge funding from the ARD2020 Biopharmaceutical program of the Région Centre Val de Loire (BioPROPHARM, CatharSIS and ETOPOCentre projects), the EU Horizon 2020 research and innovation program (MIAMi project-grant agreement N◦814645) and ANR MIACYC (ANR-20-CE43-0010).

References:

[1] B. St-Pierre, S. Besseau, M. Clastre, V. Courdavault, M. Courtois, J. Crèche, E. Ducos, T.D. de Bernonville, C. Dutilleul, G. Glévarec, N. Imbault, A. Lanoue, A. Oudin, N. Papon, O. Pichon, N. Giglioli-Guivarc'h, Deciphering the Evolution, Cell Biology and Regulation of Monoterpene Indole Alkaloids, in: N.B.T.-A. in B.R. Giglioli-Guivarc'h (Ed.), New Light Alkaloid Biosynth. Futur. Prospect., Academic Press, 2013: pp. 73–109. https://doi.org/10.1016/B978-0-12-408061-4.00003-1.

[2] T. Dugé De Bernonville, I. Carqueijeiro, A. Lanoue, F. Lafontaine, P. Sánchez Bel, F. Liesecke, K. Musset, A. Oudin, G. Glévarec, O. Pichon, S. Besseau, M. Clastre, B. St-Pierre, V. Flors, S. Maury, E. Huguet, S.E. O'Connor, V. Courdavault, Folivory elicits a strong defense reaction in *Catharanthus roseus*: Metabolomic and transcriptomic analyses reveal distinct local and systemic responses, Sci. Rep. 7 (2017). https://doi.org/10.1038/srep40453.

[3] A. Thawabteh, S. Juma, M. Bader, D. Karaman, L. Scrano, S. Bufo, R. Karaman, The Biological Activity of Natural Alkaloids against Herbivores, Cancerous Cells and Pathogens, Toxins. 11 (2019) 656. https://doi.org/10.3390/toxins11110656.

[4] E.O. Erhirhie, C. Ikegbune, A.I. Okeke, C.C. Onwuzuligbo, N.U. Madubuogwu, U.M. Chukwudulue, O.B. Okonkwo, Antimalarial herbal drugs: a review of their interactions with conventional antimalarial drugs, Clin. Phytoscience. 7 (2021) 4. https://doi.org/10.1186/s40816-020-00242-4.

[5] M. Huang, Z. Liao, X. Li, Z. Yang, X. Fan, Y. Li, Z. Zhao, S. Lang, L. Cyganek, X. Zhou, I. Akin, M. Borggrefe, I. El-Battrawy, Effects of Antiarrhythmic Drugs on hERG Gating in Human-Induced Pluripotent Stem Cell-Derived Cardiomyocytes From a Patient With Short QT Syndrome Type 1, Front. Pharmacol. 12 (2021). https://doi.org/10.3389/fphar.2021.675003.

[6] S.E. O'Connor, J.J. Maresh, Chemistry and biology of monoterpene indole alkaloid biosynthesis, Nat. Prod. Rep. 23 (2006) 532. https://doi.org/10.1039/b512615k.

[7] P. Keglevich, L. Hazai, G. Kalaus, C. Szántay, Modifications on the Basic Skeletons of Vinblastine and Vincristine, Molecules. 17 (2012) 5893–5914. https://doi.org/10.3390/molecules17055893.

[8] D.W. Carney, J.C. Lukesh, D.M. Brody, M.M. Brütsch, D.L. Boger, Ultrapotent vinblastines in which added molecular complexity further disrupts the target tubulin dimer–dimer interface, Proc. Natl. Acad. Sci. 113 (2016) 9691–9698. https://doi.org/10.1073/pnas.1611405113.

[9] V. Courdavault, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre, V. Burlat, A look inside an alkaloid multisite plant: the *Catharanthus* logistics, Curr. Opin. Plant Biol. 19 (2014) 43–50. https://doi.org/https://doi.org/10.1016/j.pbi.2014.03.010.

[10] T. Dugé de Bernonville, E. Foureau, C. Parage, A. Lanoue, M. Clastre, M.A. Londono, A. Oudin, B. Houillé, N. Papon, S. Besseau, G. Glévarec, L. Atehortùa, N. Giglioli-Guivarc'h, B. St-Pierre, V. De Luca, S.E. O'Connor, V. Courdavault, Characterization of a second secologanin synthase isoform producing both secologanin and

secoxyloganin allows enhanced de novo assembly of a *Catharanthus roseus* transcriptome, BMC Genomics. 16 (2015) 619. https://doi.org/10.1186/s12864-015-1678-y.

[11] L. Caputi, J. Franke, S.C. Farrow, K. Chung, R.M.E. Payne, T.D. Nguyen, T.T.T. Dang, I.S.T. Carqueijeiro, K. Koudounas, T.D. de Bernonville, B. Ameyaw, D.M. Jones, I.J.C. Vieira, V. Courdavault, S.E. O'Connor, Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle, Science. 360 (2018) 1235–1239. https://doi.org/10.1126/science.aat4100.

[12] S. Brown, M. Clastre, V. Courdavault, S.E. O'Connor, De novo production of the plant-derived alkaloid strictosidine in yeast, Proc. Natl. Acad. Sci. 112 (2015) 3205–3210. https://doi.org/10.1073/pnas.1423555112.

[13] Y. Qu, M.L.A.E. Easson, J. Froese, R. Simionescu, T. Hudlicky, V. De Luca, Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast, Proc. Natl. Acad. Sci. 112 (2015) 6224–6229. https://doi.org/10.1073/pnas.1501821112.

[14] I. Carqueijeiro, T. Dugé de Bernonville, A. Lanoue, T.-T. Dang, C.N. Teijaro, C. Paetz, K. Billet, A. Mosquera, A. Oudin, S. Besseau, N. Papon, G. Glévarec, L. Atehortùa, M. Clastre, N. Giglioli-Guivarc'h, B. Schneider, B. St-Pierre, R.B. Andrade, S.E. O'Connor, V. Courdavault, A BAHD acyltransferase catalyzing 19- *O* -acetylation of tabersonine derivatives in roots of *Catharanthus roseus* enables combinatorial synthesis of monoterpene indole alkaloids, Plant J. 94 (2018) 469–484. https://doi.org/10.1111/tpj.13868.

[15] P. Lemos Cruz, N. Kulagina, G. Guirimand, J.-O. De Craene, S. Besseau, A. Lanoue, A. Oudin, N. Giglioli-Guivarc'h, N. Papon, M. Clastre, V. Courdavault, Optimization of Tabersonine Methoxylation to Increase Vindoline Precursor Synthesis in Yeast Cell Factories, Molecules. 26 (2021) 3596. https://doi.org/10.3390/molecules26123596.

[16] N. Kulagina, G. Guirimand, C. Melin, P. Lemos-Cruz, I. Carqueijeiro, J. De Craene, A. Oudin, V. Heredia, K. Koudounas, M. Unlubayir, A. Lanoue, N. Imbault, B. St-Pierre, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, J. Marc, S. Besseau, V. Courdavault, Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories, Microb. Biotechnol. (2021) 1751-7915.13898. https://doi.org/10.1111/1751-7915.13898.

[17] V. De Luca, V. Salim, A. Thamm, S.A. Masada, F. Yu, Making iridoids/secoiridoids and monoterpenoid indole alkaloids: progress on pathway elucidation, Curr. Opin. Plant Biol. 19 (2014) 35–42. https://doi.org/10.1016/j.pbi.2014.03.006.

[18] R. Sadre, M. Magallanes-Lundback, S. Pradhan, V. Salim, A. Mesberg, A.D. Jones, D. DellaPenna, Metabolite Diversity in Alkaloid Biosynthesis: A Multilane (Diastereomer) Highway for Camptothecin Synthesis in *Camptotheca acuminata*, Plant Cell. 28 (2016) 1926–1944. https://doi.org/10.1105/tpc.16.00193.

[19] B.R. Lichman, The scaffold-forming steps of plant alkaloid biosynthesis, Nat. Prod. Rep. 38 (2021) 103–129. https://doi.org/10.1039/D0NP00031K.

[20] P.F. Rosales, G.S. Bordin, A.E. Gower, S. Moura, Indole alkaloids: 2012 until now, highlighting the new chemical structures and biological activities, Fitoterapia. 143 (2020) 104558. https://doi.org/10.1016/j.fitote.2020.104558.

[21] M. Sottomayor, A. Ros Barceló, Peroxidase from *Catharanthus roseus* (L.) G. Don and the biosynthesis of alpha-3',4'-anhydrovinblastine: a specific role for a multifunctional enzyme, Protoplasma. 222 (2003) 97–105. https://doi.org/10.1007/s00709-003-0003-9.

[22] H. Ishikawa, D.A. Colby, S. Seto, P. Va, A. Tam, H. Kakei, T.J. Rayl, I. Hwang, D.L. Boger, Total Synthesis of Vinblastine, Vincristine, Related Natural Products, and Key Structural Analogues, J. Am. Chem. Soc. 131 (2009) 4904–4916. https://doi.org/10.1021/ja809842b.

[23] P. Westekemper, U. Wieczorek, F. Gueritte, N. Langlois, Y. Langlois, P. Potier, M. Zenk, Radioimmunoassay for the Determination of the Indole Alkaloid Vindoline in *Catharanthus*, Planta Med. 39 (1980) 24–37. https://doi.org/10.1055/s-2008-1074900.

[24] B. St-Pierre, V. De Luca, A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus*, Plant Physiol. 109 (1995) 131–139. https://doi.org/10.1104/pp.109.1.131.

[25] G. Schröder, E. Unterbusch, M. Kaltenbach, J. Schmidt, D. Strack, V. De Luca, J. Schröder, Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase, FEBS Lett. 458 (1999) 97–102. https://doi.org/10.1016/S0014-5793(99)01138-2.

[26] L.-A. Giddings, D.K. Liscombe, J.P. Hamilton, K.L. Childs, D. DellaPenna, C.R. Buell, S.E. O'Connor, A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*, J. Biol. Chem. 286 (2011) 16751–16757. https://doi.org/10.1074/jbc.M111.225383.

[27] I. Carqueijeiro, S. Brown, K. Chung, T.-T. Dang, M. Walia, S. Besseau, T. Dugé de Bernonville, A. Oudin, A. Lanoue, K. Billet, T. Munsch, K. Koudounas, C. Melin, C. Godon, B. Razafimandimby, J.-O. de Craene, G. Glévarec, J. Marc, N. Giglioli-Guivarc'h, M. Clastre, B. St-Pierre, N. Papon, R.B. Andrade, S.E. O'Connor, V. Courdavault, Two Tabersonine 6,7-Epoxidases Initiate Lochnericine-Derived Alkaloid Biosynthesis in *Catharanthus roseus*, Plant Physiol. 177 (2018) 1473–1486. https://doi.org/10.1104/pp.18.00549.

[28] D. Williams, Y. Qu, R. Simionescu, V. De Luca, The assembly of (+)-vincadifformine- and (−)-tabersonine-derived monoterpenoid indole alkaloids in *Catharanthus roseus* involves separate branch pathways, Plant J. 99 (2019) 626–636. https://doi.org/10.1111/tpj.14346.

[29] D. Levac, J. Murata, W.S. Kim, V. De Luca, Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O- methyltransferase, Plant J. 53 (2008) 225–236. https://doi.org/10.1111/j.1365-313X.2007.03337.x.

[30] V. De Luca, J.A. Fernandez, D. Campbell, W.G.W. Kurz, Developmental Regulation of Enzymes of Indole Alkaloid Biosynthesis in *Catharanthus roseus*, Plant Physiol. 86

(1988) 447–450. https://doi.org/10.1104/pp.86.2.447.

[31] G. Schröder, E. Wehinger, J. Schröder, Predicting the substrates of cloned plant O-methyltransferases, Phytochemistry. 59 (2002) 1–8. https://doi.org/10.1016/S0031-9422(01)00421-6.

[32] S. Cacace, G. Schröder, E. Wehinger, D. Strack, J. Schmidt, J. Schröder, A flavonol O-methyltransferase from *Catharanthus roseus* performing two sequential methylations, Phytochemistry. 62 (2003) 127–137. https://doi.org/https://doi.org/10.1016/S0031-9422(02)00483-1.

[33] G. Schröder, E. Wehinger, R. Lukačin, F. Wellmann, W. Seefelder, W. Schwab, J. Schröder, Flavonoid methylation: a novel 4′-O-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases, Phytochemistry. 65 (2004) 1085–1094. https://doi.org/10.1016/j.phytochem.2004.02.010.

[34] G. Guirimand, A. Guihur, P. Poutrain, F. Héricourt, S. Mahroug, B. St-Pierre, V. Burlat, V. Courdavault, Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*, J. Plant Physiol. 168 (2011) 549–557. https://doi.org/10.1016/j.jplph.2010.08.018.

[35] F. Kellner, F. Geu-Flores, N.H. Sherden, S. Brown, E. Foureau, V. Courdavault, S.E. O'Connor, Discovery of a P450-catalyzed step in vindoline biosynthesis: a link between the aspidosperma and eburnamine alkaloids, Chem. Commun. 51 (2015) 7626–7628. https://doi.org/10.1039/C5CC01309G.

[36] D.K. Liscombe, A.R. Usera, S.E. O'Connor, Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis, Proc. Natl. Acad. Sci. 107 (2010) 18793–18798. https://doi.org/10.1073/pnas.1009003107.

[37] F. Vazquez-Flota, E. De Carolis, A.M. Alarco, V. De Luca, Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in *Catharanthus roseus* (L.) G. Don, Plant Mol. Biol. 34 (1997) 935–948. https://doi.org/10.1023/A:1005894001516.

[38] E. De Carolis, V. De Luca, A novel 2-oxoglutarate-dependent dioxygenase involved in vindoline biosynthesis: characterization, purification and kinetic properties, Plant Cell. Tissue Organ Cult. 38 (1994) 281–287. https://doi.org/10.1007/BF00033888.

[39] I. Carqueijeiro, K. Koudounas, T. Dugé de Bernonville, L.J. Sepúlveda, A. Mosquera, D.P. Bomzan, A. Oudin, A. Lanoue, S. Besseau, P. Lemos Cruz, N. Kulagina, E.A. Stander, S. Eymieux, J. Burlaud-Gaillard, E. Blanchard, M. Clastre, L. Atehortùa, B. St-Pierre, N. Giglioli-Guivarc'h, N. Papon, D.A. Nagegowda, S.E. O'Connor, V. Courdavault, Alternative splicing creates a pseudo-strictosidine β-d-glucosidase modulating alkaloid synthesis in *Catharanthus roseus*, Plant Physiol. 185 (2021) 836–856. https://doi.org/10.1093/plphys/kiaa075.

[40] A. Stavrinides, E.C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S.E. O'Connor, Unlocking the diversity of alkaloids in *Catharanthus roseus*: Nuclear localization suggests metabolic channeling in secondary metabolism, Chem. Biol. 22

(2015) 336–341. https://doi.org/10.1016/j.chembiol.2015.02.006.

[41] A. Stavrinides, E.C. Tatsis, L. Caputi, E. Foureau, C.E.M. Stevenson, D.M. Lawson, V. Courdavault, S.E. O'Connor, Structural investigation of heteroyohimbine alkaloid synthesis reveals active site elements that control stereoselectivity, Nat. Commun. 7 (2016) 12116. https://doi.org/10.1038/ncomms12116.

[42] S. Besseau, F. Kellner, A. Lanoue, A.M.K. Thamm, V. Salim, B. Schneider, F. Geu-Flores, R. Höfer, G. Guirimand, A. Guihur, A. Oudin, G. Glevarec, E. Foureau, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre, D. Werck-Reichhart, V. Burlat, V. De Luca, S.E. O'Connor, V. Courdavault, A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*, Plant Physiol. 163 (2013) 1792–1803. https://doi.org/10.1104/pp.113.222828.

[43] M.R. Park, X. Chen, D.E. Lang, K.K.S. Ng, P.J. Facchini, Heterodimeric O -methyltransferases involved in the biosynthesis of noscapine in opium poppy, Plant J. 95 (2018) 252–267. https://doi.org/10.1111/tpj.13947.

[44] I.M. Menéndez-Perdomo, P.J. Facchini, Isolation and characterization of two O-methyltransferases involved in benzylisoquinoline alkaloid biosynthesis in sacred lotus (*Nelumbo nucifera*), J. Biol. Chem. 295 (2020) 1598–1612. https://doi.org/10.1074/jbc.RA119.011547.

[45] J. Franke, J. Kim, J.P. Hamilton, D. Zhao, G.M. Pham, K. Wiegert-Rininger, E. Crisovan, L. Newton, B. Vaillancourt, E. Tatsis, C.R. Buell, S.E. O'Connor, Gene Discovery in *Gelsemium* Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis, ChemBioChem. 20 (2019) 83–87. https://doi.org/10.1002/cbic.201800592.

[46] I. Carqueijeiro, E. Masini, E. Foureau, L.J. Sepúlveda, E. Marais, A. Lanoue, S. Besseau, N. Papon, M. Clastre, T. Dugé de Bernonville, G. Glévarec, L. Atehortùa, A. Oudin, V. Courdavault, Virus-induced gene silencing in *Catharanthus roseus* by biolistic inoculation of tobacco rattle virus vectors, Plant Biol. 17 (2015) 1242–1246. https://doi.org/10.1111/plb.12380.

[47] G. Guirimand, V. Burlat, A. Oudin, A. Lanoue, B. St-Pierre, V. Courdavault, Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase, Plant Cell Rep. 28 (2009) 1215–1234. https://doi.org/10.1007/s00299-009-0722-2.

[48] J. Murata, J. Roepke, H. Gordon, V. De Luca, The Leaf Epidermome of *Catharanthus roseus* Reveals Its Biochemical Specialization, Plant Cell. 20 (2008) 524–542. https://doi.org/10.1105/tpc.107.056630.

[49] D. Pompon, B. Louerat, A. Bronine, P. Urban, Yeast expression of animal and plant P450s in optimized redox environments, in: Methods Enzymol., Academic Press Inc., 1996: pp. 51–64. https://doi.org/10.1016/S0076-6879(96)72008-6.

[50] C. Parage, E. Foureau, F. Kellner, V. Burlat, S. Mahroug, A. Lanoue, T.D. de Bernonville, M.A. Londono, I. Carqueijeiro, A. Oudin, S. Besseau, N. Papon, G.

Glévarec, L. Atehortùa, N. Giglioli-Guivarc'h, B. St-Pierre, M. Clastre, S.E. O'Connor, V. Courdavault, Class II cytochrome P450 reductase governs the biosynthesis of alkaloids, Plant Physiol. 172 (2016) 1563–1577. https://doi.org/10.1104/pp.16.00801.

[51] M. Bradford, A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding, Anal. Biochem. 72 (1976) 248–254. https://doi.org/10.1006/abio.1976.9999.

[52] U.K. Laemmli, Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4, Nature. 227 (1970) 680–685. https://doi.org/10.1038/227680a0.

[53] G. Guirimand, V. Courdavault, A. Lanoue, S. Mahroug, A. Guihur, N. Blanc, N. Giglioli-Guivarc'h, B. St-Pierre, V. Burlat, Strictosidine activation in Apocynaceae: Towards a "nuclear time bomb"?, BMC Plant Biol. 10 (2010) 1–20. https://doi.org/10.1186/1471-2229-10-182.

[54] K. Koudounas, M. Thomopoulou, C. Michaelidis, E. Zevgiti, G. Papakostas, P. Tserou, G. Daras, P. Hatzopoulos, The C-Domain of Oleuropein β -Glucosidase Assists in Protein Folding and Sequesters the Enzyme in Nucleus, Plant Physiol. 174 (2017) 1371–1383. https://doi.org/10.1104/pp.17.00512.

[55] E.A. Stander, L.J. Sepúlveda, T. Dugé de Bernonville, I. Carqueijeiro, K. Koudounas, P. Lemos Cruz, S. Besseau, A. Lanoue, N. Papon, N. Giglioli-Guivarc'h, R. Dirks, S.E. O'Connor, L. Atehortùa, A. Oudin, V. Courdavault, Identifying Genes Involved in Alkaloid Biosynthesis in *Vinca minor* through Transcriptomics and Gene Co-Expression Analysis, Biomolecules. 10 (2020) 1595. https://doi.org/10.3390/biom10121595.

[56] G. Guirimand, V. Burlat, A. Oudin, A. Lanoue, B. St-Pierre, V. Courdavault, Optimization of the transient transformation of *Catharanthus roseus* cells by particle bombardment and its application to the subcellular localization of hydroxymethylbutenyl 4-diphosphate synthase and geraniol 10-hydroxylase, Plant Cell Rep. 28 (2009) 1215–1234. https://doi.org/10.1007/s00299-009-0722-2.

[57] R. Waadt, L.K. Schmidt, M. Lohse, K. Hashimoto, R. Bock, J. Kudla, Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta, Plant J. 56 (2008) 505–516. https://doi.org/10.1111/j.1365-313X.2008.03612.x.

[58] E. Foureau, I. Carqueijeiro, T. Dugé de Bernonville, C. Melin, F. Lafontaine, S. Besseau, A. Lanoue, N. Papon, A. Oudin, G. Glévarec, M. Clastre, B. St-Pierre, N. Giglioli-Guivarc'h, V. Courdavault, Prequels to Synthetic Biology: From Candidate Gene Identification and Validation to Enzyme Subcellular Localization in Plant and Yeast Cells, Methods Enzymol. 576 (2016) 167–206. https://doi.org/10.1016/bs.mie.2016.02.013.

[59] A.K. Singh, V. Dwivedi, A. Rai, S. Pal, S.G.E. Reddy, D.K.V. Rao, A.K. Shasany, D.A. Nagegowda, Virus-induced gene silencing of *Withania somnifera* squalene synthase negatively regulates sterol and defence-related genes resulting in reduced withanolides and biotic stress tolerance, Plant Biotechnol. J. 13 (2015) 1287–1299. https://doi.org/10.1111/pbi.12347.

[60] P.L. Cruz, M.I. Restrepo, T.D. de Bernonville, A. Oudin, T. Munsch, A. Lanoue, S. Besseau, L. Atehortùa, N. Giglioli-Guivarc'h, N. Papon, M. Clastre, I. Carqueijeiro, V. Courdavault, A Biolistic-Mediated Virus-Induced Gene Silencing in Apocynaceae to Map Biosynthetic Pathways of Alkaloids, in: 2020: pp. 93–110. https://doi.org/10.1007/978-1-0716-0751-0_8.

[61] B. Ahmad, A. Banerjee, H. Tiwari, S. Jana, S. Bose, S. Chakrabarti, Structural and functional characterization of the Vindoline biosynthesis pathway enzymes of *Catharanthus roseus*, J. Mol. Model. 24 (2018) 1–14. https://doi.org/10.1007/s00894-018-3590-2.

[62] V. Salim, A.D. Jones, D. DellaPenna, D. Jones, D. DellaPenna, *Camptotheca acuminata* 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism, Plant J. 95 (2018) 112–125. https://doi.org/10.1111/tpj.13936.

[63] S.C. Farrow, M.O. Kamileen, J. Meades, B. Ameyaw, Y. Xiao, S.E. O'Connor, Cytochrome P450 and O-methyltransferase catalyze the final steps in the biosynthesis of the anti-addictive alkaloid ibogaine from *Tabernanthe iboga*, J. Biol. Chem. 293 (2018) 13821–13833. https://doi.org/10.1074/jbc.RA118.004060.

[64] T. Nomura, T.M. Kutchan, Three New O-Methyltransferases Are Sufficient for All O-Methylation Reactions of Ipecac Alkaloid Biosynthesis in Root Culture of *Psychotria ipecacuanha*, J. Biol. Chem. 285 (2010) 7722–7738. https://doi.org/10.1074/jbc.M109.086157.

[65] A. Ounaroon, G. Decker, J. Schmidt, F. Lottspeich, T.M. Kutchan, (R,S)-Reticuline 7-O -methyltransferase and (R,S)-norcoclaurine 6- O -methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy, Plant J. 36 (2003) 808–819. https://doi.org/10.1046/j.1365-313X.2003.01928.x.

[66] S. Pienkny, W. Brandt, J. Schmidt, R. Kramell, J. Ziegler, Functional characterization of a novel benzylisoquinoline O -methyltransferase suggests its involvement in papaverine biosynthesis in opium poppy (*Papaver somniferum L*), Plant J. 60 (2009) 56–67. https://doi.org/10.1111/j.1365-313X.2009.03937.x.

[67] D.K. Liscombe, S.E. O'Connor, A virus-induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*, Phytochemistry. 72 (2011) 1969–1977. https://doi.org/10.1016/j.phytochem.2011.07.001.

[68] E.C. Tatsis, I. Carqueijeiro, T. Dugé De Bernonville, J. Franke, T.T.T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A.K. Stavrinides, M. Clastre, V. Courdavault, S.E. O'Connor, A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate, Nat. Commun. 8 (2017) 1–9. https://doi.org/10.1038/s41467-017-00154-x.

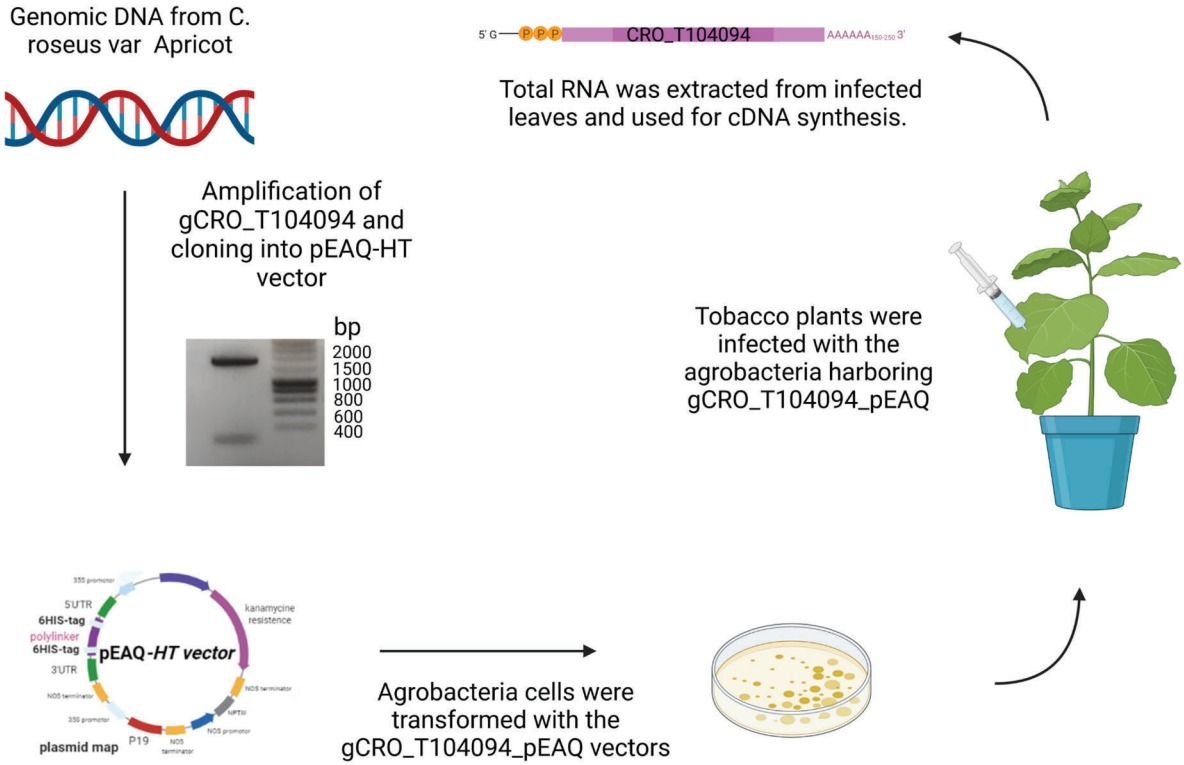
[69] V. Salim, F. Yu, J. Altarejos, V. De Luca, Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis, Plant J. 76 (2013) 754–765. https://doi.org/10.1111/tpj.12330.

[70] F. Ferreres, D.M. Pereira, P. Valentão, P.B. Andrade, R.M. Seabra, M. Sottomayor, New Phenolic Compounds and Antioxidant Potential of *Catharanthus roseus*, J. Agric. Food Chem. 56 (2008) 9967–9974. https://doi.org/10.1021/jf8022723.

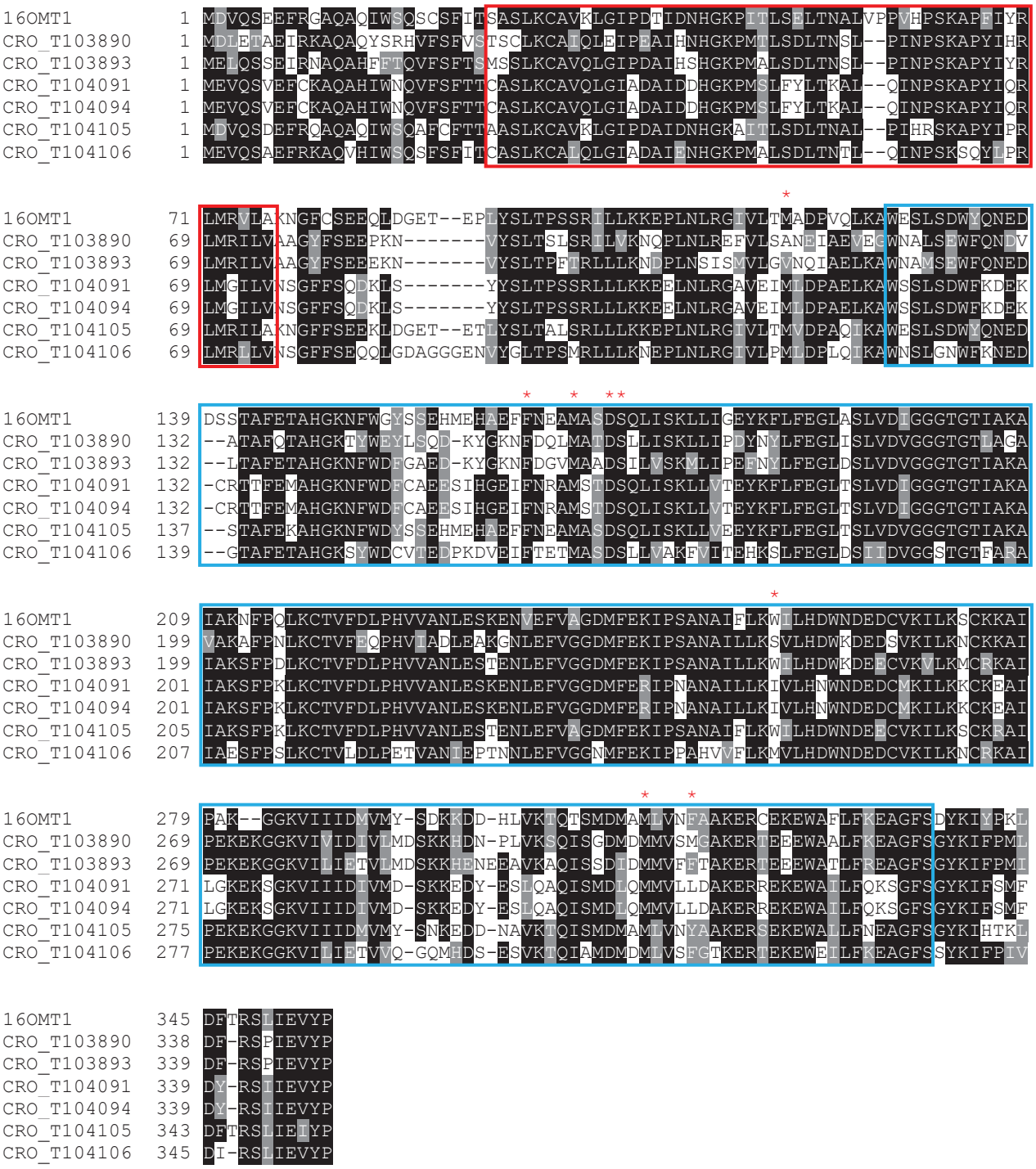
[71] J. Liu, Y. Liu, Y. Wang, Z.-H. Zhang, Y.-G. Zu, T. Efferth, Z.-H. Tang, The Combined Effects of Ethylene and MeJA on Metabolic Profiling of Phenolic Compounds in *Catharanthus roseus* Revealed by Metabolomics Analysis, Front. Physiol. 7 (2016). https://doi.org/10.3389/fphys.2016.00217.

[72] I. Carqueijeiro, H. Noronha, P. Duarte, H. Gerós, M. Sottomayor, Vacuolar Transport of the Medicinal Alkaloids from *Catharanthus roseus* Is Mediated by a Proton-Driven Antiport, Plant Physiol. 162 (2013) 1486–1496. https://doi.org/10.1104/pp.113.220558.

Supplementary data

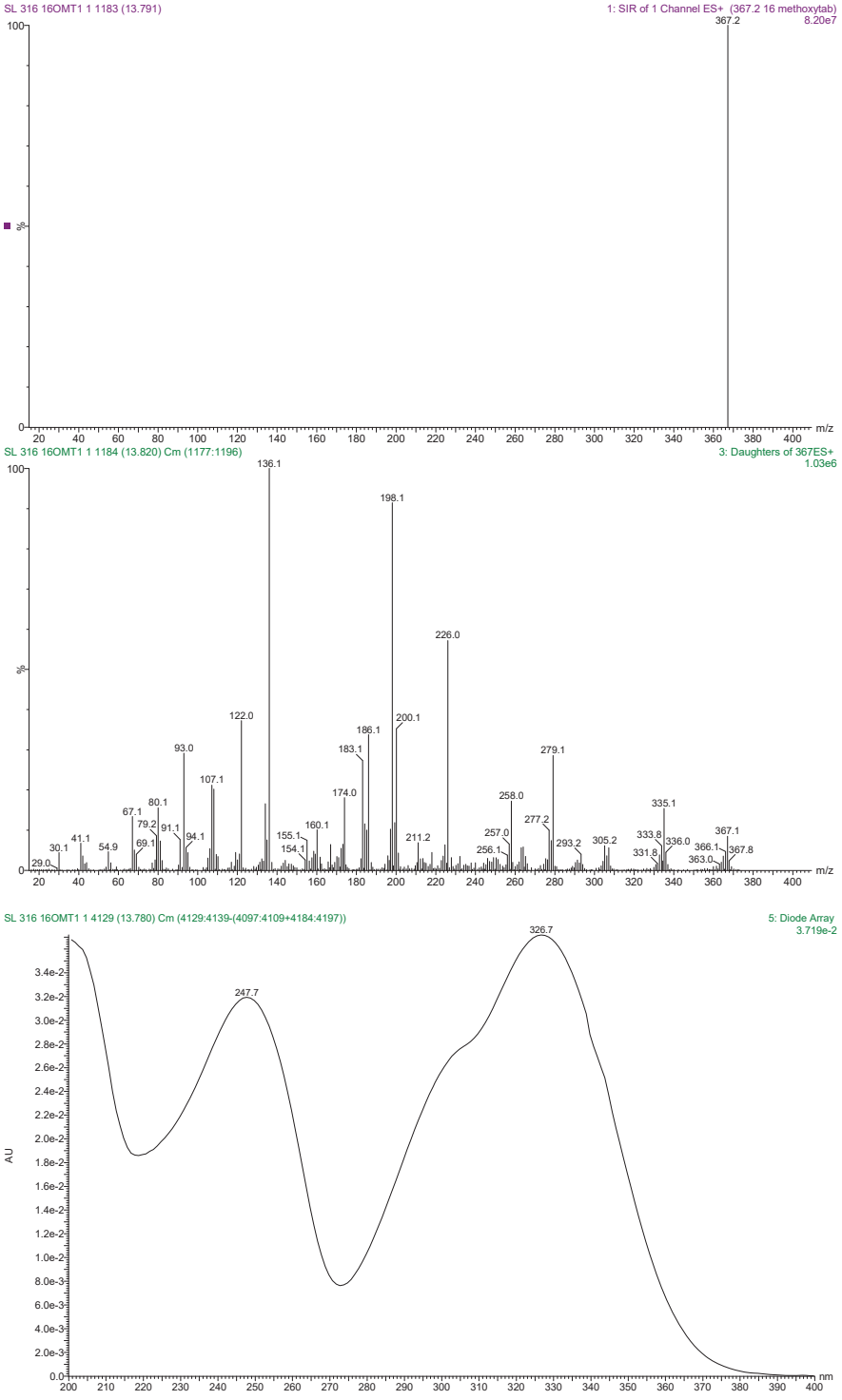


Supplemental Figure S1. In consequence to the lack of amplification of CRO_T104094 from C. *roseus* cDNA, this candidate was amplified from genomic DNA from *C. roseus.* The procedure for introns removal was done according to the figure. The genomic sequence of CRO_T104094 was cloned into a pEAQ-HT vector which was subsequently introduced in Agrobacteria. Tobacco plants were infected with the Agrobacteria harboring the OMT3-pEAQ plasmids. After 5 days, the infected leaves were collected and total RNA was extracted. The mRNAs were transcribed to cDNA that were used for Phusion PCR amplification of the CRO_T104094 candidate. "Created with BioRender.com"

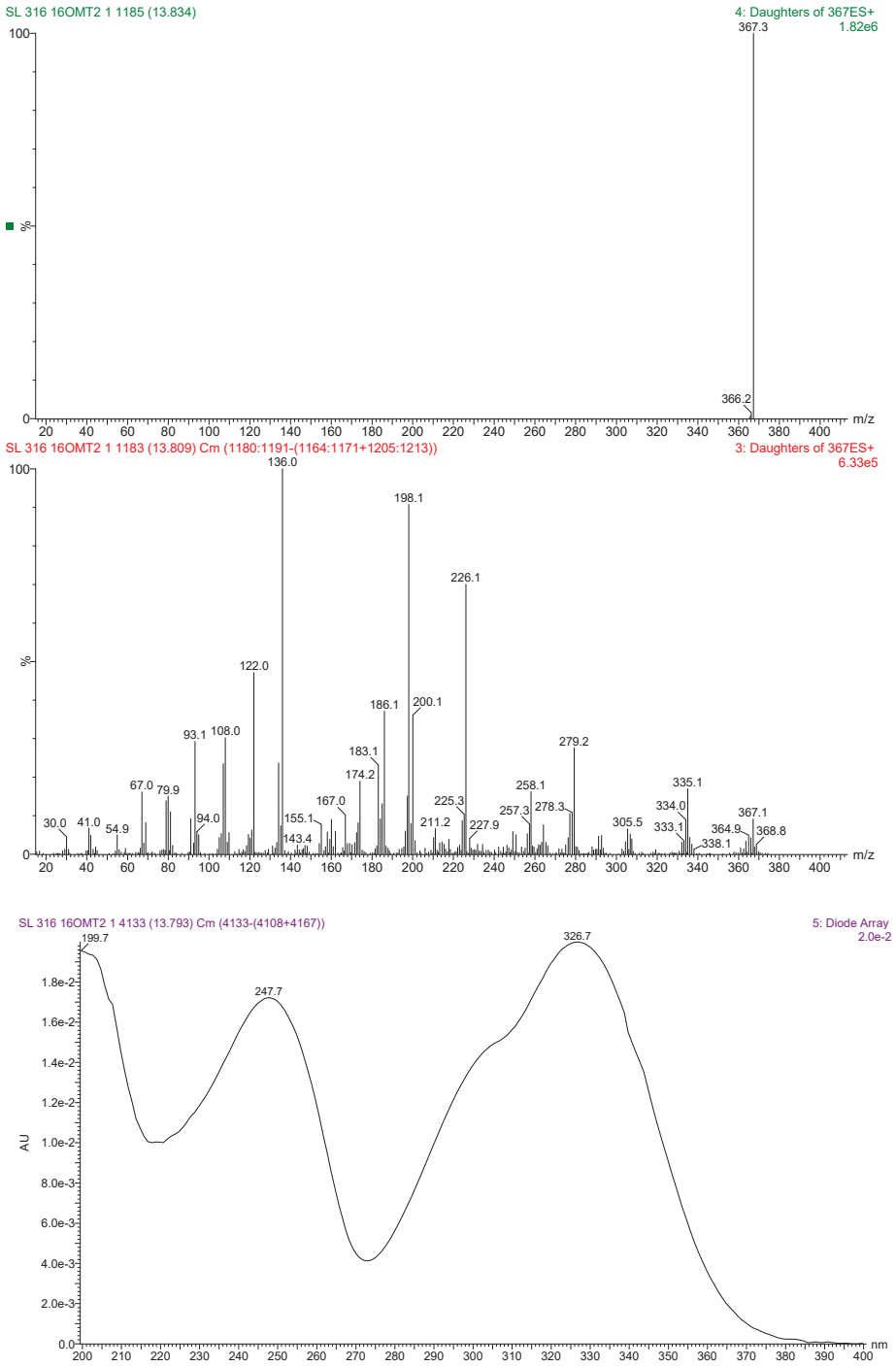


Supplemental Figure 2. Alignment of the amino-acid sequences of 16OMT1 with flavonoid *O*-methyltransferases (CRO_T103980, CRO_T103093, T104091 and CRO_T104106) and 16OMT candidates (CRO_T104094 and CRO_T104105). Sequence identity and similarity are highlighted by black and grey shading, respectively. The dimerization domain is highlighted in red and the *O*-methyltransferase domain is highlighted in blue. Red stars indicate residues predicted to be involved in interaction with 16-hydroxytabersonine according to Ahmad et al., (2018).

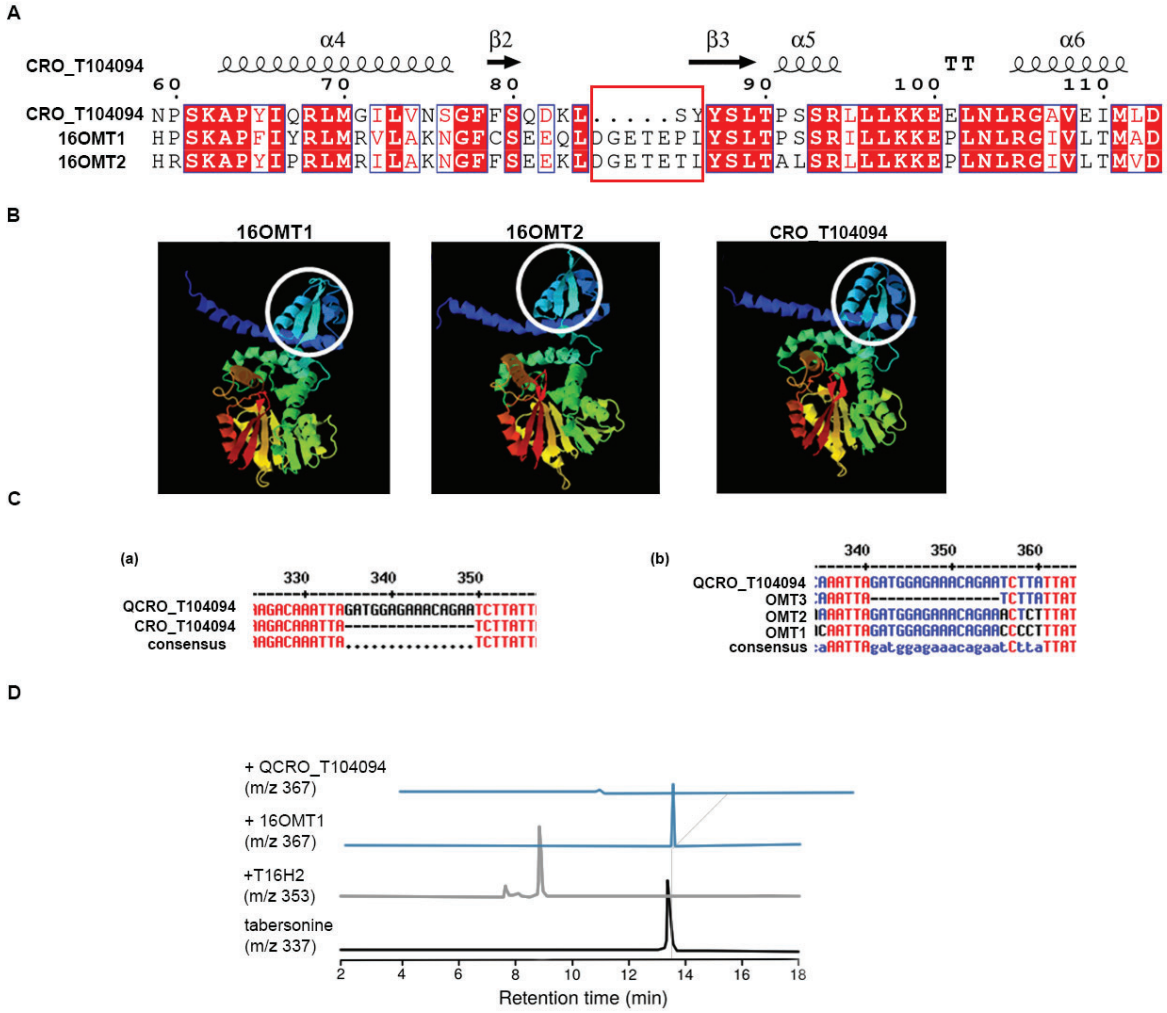
A


SL 316 16OMT1 1 1183 (13.791)
1: SIR of 1 Channel ES+ (367.2 16 methoxytab)
367.2
8.20e7
100
%
0
20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400
m/z
SL 316 16OMT1 1 1184 (13.820) Cm (1177:1196)
3: Daughters of 367ES+
1.03e6
136.1
198.1
226.0
122.0
200.1
186.1
183.1
93.0
279.1
107.1
174.0
258.0
335.1
80.1
67.1
277.2
79.2
91.1
160.1
257.0
367.1
333.8
94.1
155.1
211.2
336.0
41.1
69.1
154.1
293.2
305.2
366.1
30.1
54.9
256.1
331.8
367.8
29.0
363.0
SL 316 16OMT1 1 4129 (13.780) Cm (4129:4139-(4097:4109+4184:4197))
5: Diode Array
3.719e-2
326.7
247.7
AU
3.4e-2
3.2e-2
3.0e-2
2.8e-2
2.6e-2
2.4e-2
2.2e-2
2.0e-2
1.8e-2
1.6e-2
1.4e-2
1.2e-2
1.0e-2
8.0e-3
6.0e-3
4.0e-3
2.0e-3
0.0
200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400
nm

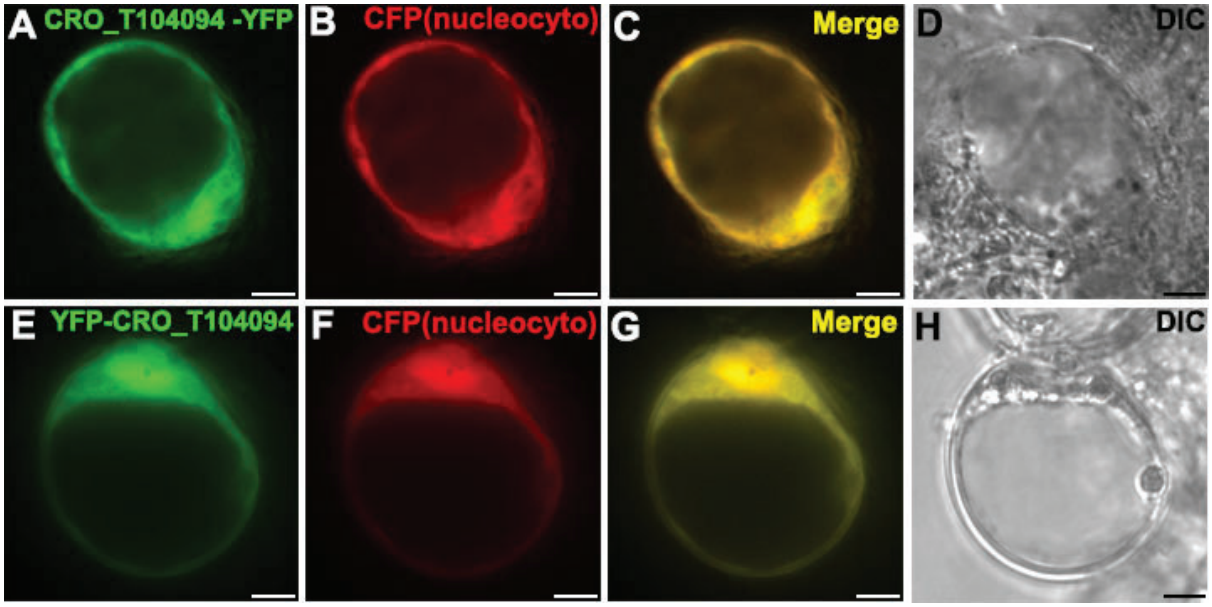
B



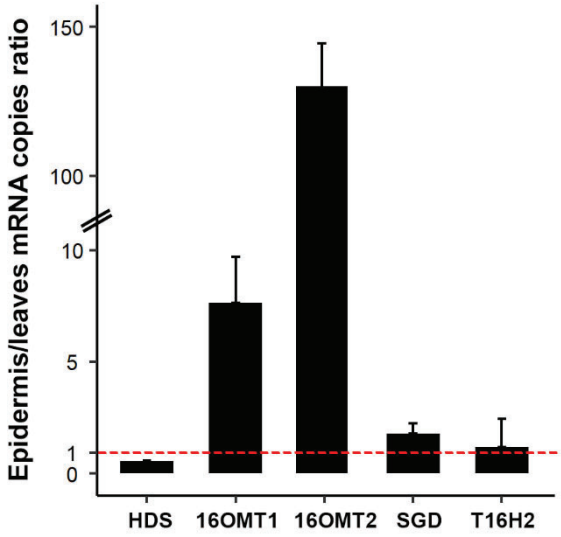
Supplemental Figure S3. MS/MS and UV spectra of the 16-methoxytabersonine produced by 16OMT1 (A) and 16OMT2 (B).



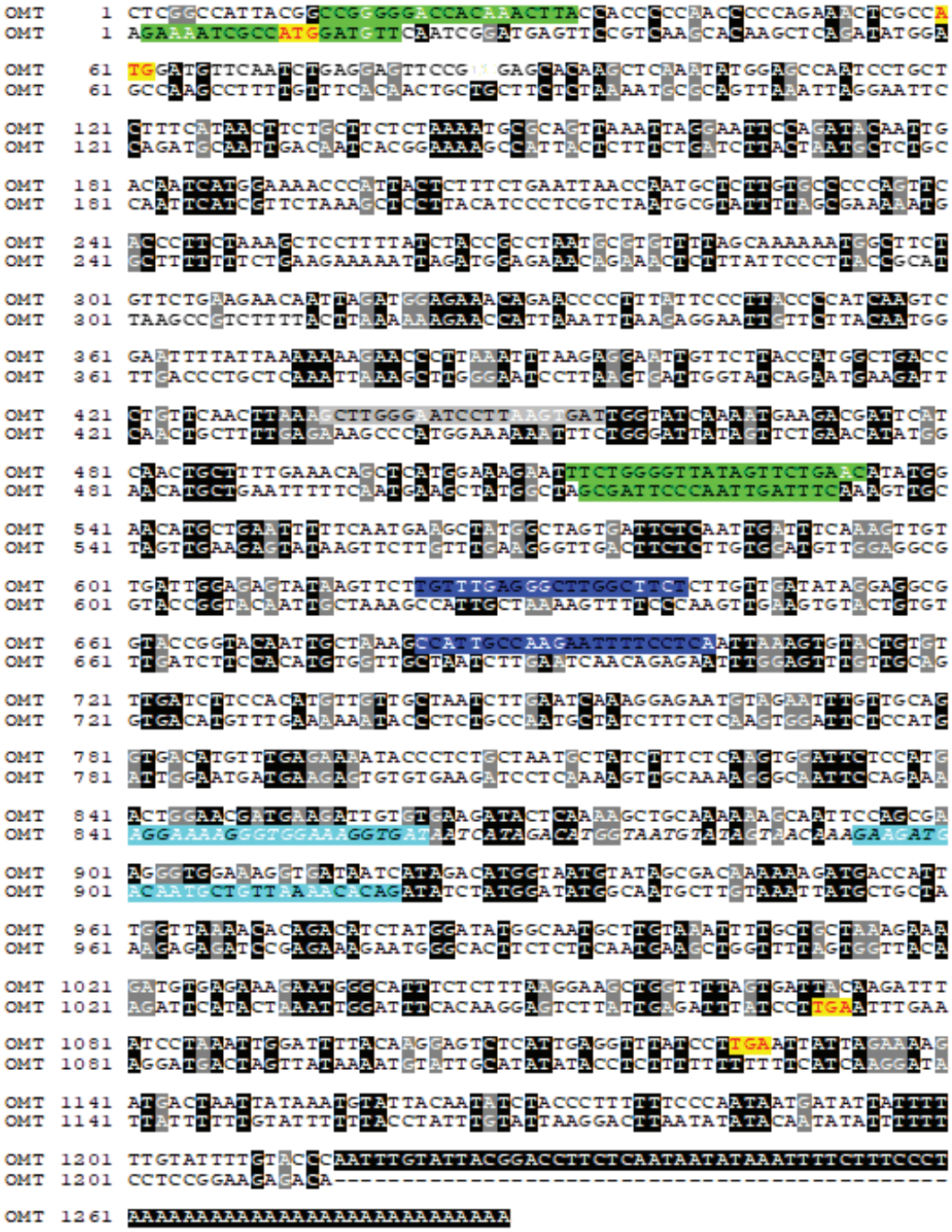
Supplemental Figure S4. (A) Sequence alignment between the two 16OMT isoforms and the CRO_T104094 candidate. On the upper part of the alignment, the predicted structure for CRO_T104094 based on the crystal structure of the isoflavanone 4'-*O*-methyltransferase. The red square shows a gap from the position 84 and the position 87, which is highly conserved on the sequence of 16OMT1 and 16OMT2. This gap is possibly located on the domain responsible for the β2 beta sheet. (B) Predicted tertiary structure of 16OMT1, 16OMT2 and CRO_T104094 as computed by the Phyre2 server [59]. The tertiary structure put in evidence a possible shortening of the beta sheet structure of CRO_T104094 (white circle). (C) (a) Constructed chimeric CRO_T104094 (QCRO_T104094) in pESC-LEU with the inserted 15 missing amino acids. (b) Alignment of the chimeric QCRO_T104094 with the 16OMT1 and 16OMT2, confirming the correction of the gap. (D) The UPLC-MS/MS analysis of the products from the yeast strain (T16H2 + QCRO_T104094) after the galactose induction and tabersonine feeding did not show the presence of 16-metoxytabersonine, as presented by the chromatograms above, thus no activity was restored for this protein.



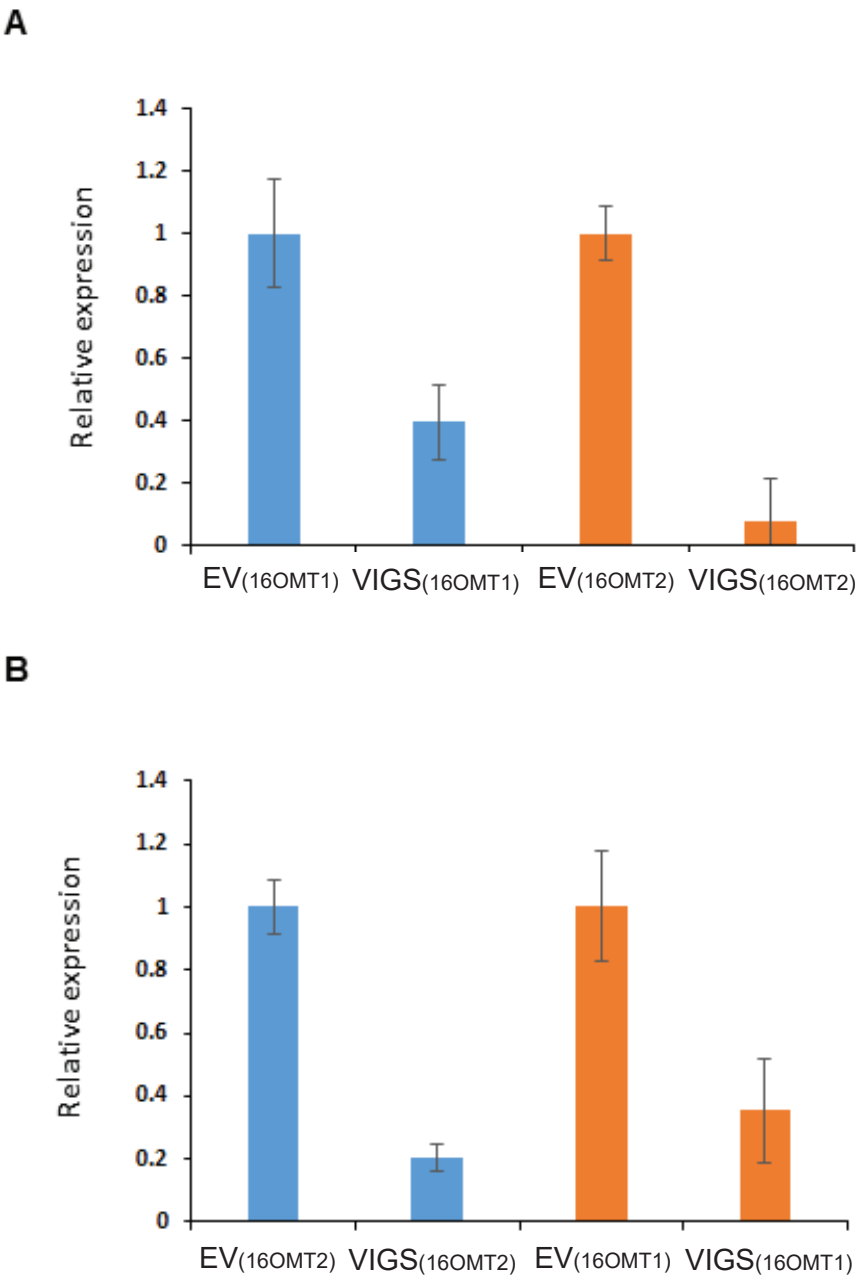
Supplemental Figure S5 - Subcellular localization of CRO_T104094 in *C. roseus* cells. Cells were transformed transiently with CRO_T104094-YFP (A–D), YFP-CRO_T104094 (E-H) expressing vectors in combination with CFP nuclear-cytosolic marker as mentioned on the images of the second column. Co-localization of the two fluorescence signals appeared on the merged image (C, G). The morphology is observed with differential interference contrast (DIC). Bar: 10 µm.



Supplemental Figure S6. 16OMT1 and 16OMT2 transcripts are mostly found in epidermis of *C. roseus* leaves. Relative expression of hydroxymethylbutenyl 4-diphosphate synthase (HDS), 16OMT1, 16OMT2, Strictosidine β-D-glucosidase (SGD) and T16H2 in epidermis enriched fractions of *C. roseus* leaves. Epidermis enriched transcript fractions were generated by a carborundum abrasion. Transcript copy numbers were normalized using CrRPS9 and expressed relatively to the amount of transcript measured in the whole leaf fraction (Epidermis/leaves ratio). The red line marks the ratio 1: Below this line, a bigger number of transcripts are found in leaves in relation to epidermis. While above this line, the transcripts are mostly found in epidermis in relation to leaves amount. Error bars correspond to the standard deviation of biological replicates (n = 3).



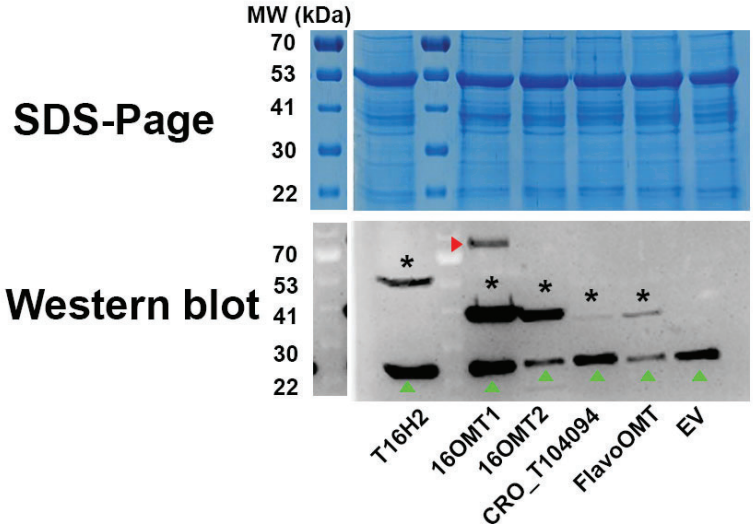
Supplemental Figure S7- Multiple sequence alignment of 16OMT nucleotide sequences (16OMT1 and 16OMT2 correspond to upper and lower line, respectively) showing the location of VIGS primers and qRT primers. Green highlight- VIGS Primers, yellow highlight- start and stop codon, blue highlights- qRT primers.



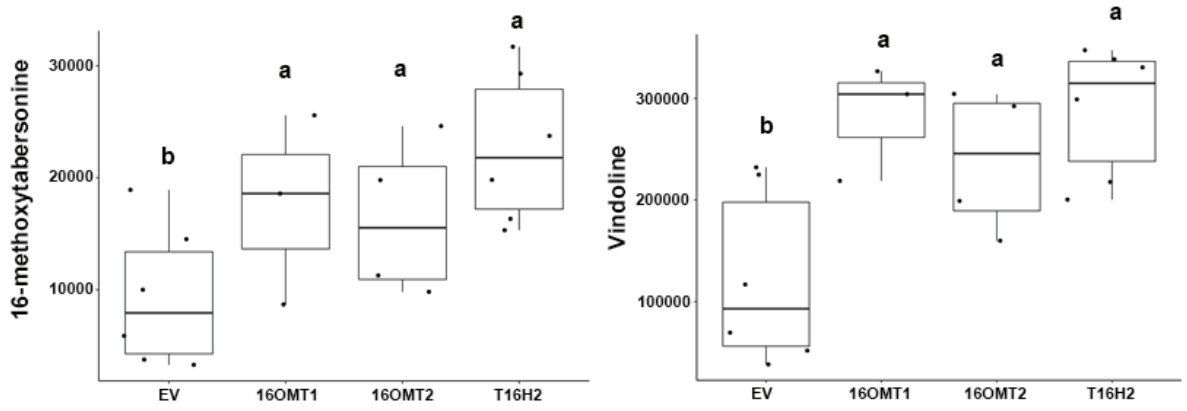
Supplemental Figure S8- Relative expression of 16OMT1 and 16OMT2 by target VIGS silencing in *C. roseus.* In this first VIGS experiment, the silencing of each isoform was aimed by selecting a specific region for each gene (Supplementary Figure S4). Expression of 16OMT1 (A) and 16OMT2 (B) in pTRV2::16OMT1 and pTRV2::16OMT2 silenced plants. qPCR shows that both isoforms had a significant reduction on expression when compared to the plants transformed with the empty vector, characterizing a cross-silencing. Error bars correspond to the standard deviation of biological replicates (n = 3).



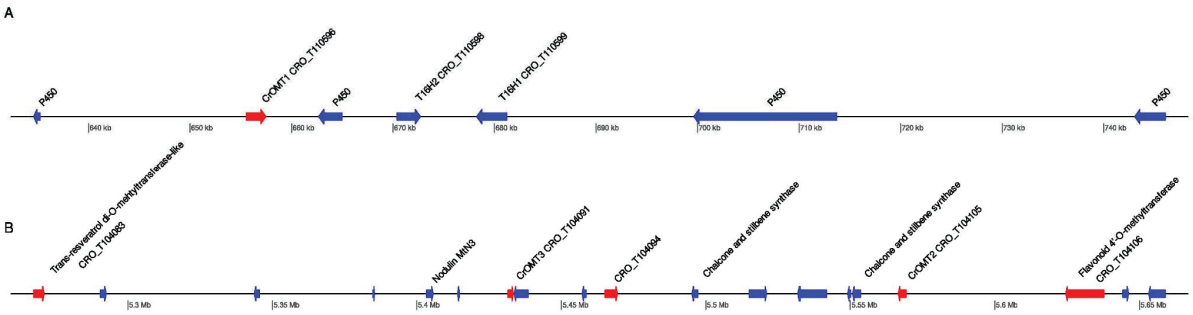
Supplemental Figure 9. Alignment of the nucleic acid sequences of 16OMT1 and 16OMT2 used for co-silencing. Sequence identity and similarity are highlighted by black and grey shading, respectively.



Supplemental Figure S10. SDS-PAGE (Coomassie blue or stain free) and Western Blot analysis of crude protein extracts (15 µg per lane) from *C. roseus* leaves agroinfiltrated with pEAQ-HT:GFP-6His and pEAQ-HT:T16H2-6His, or pEAQ-HT:16OMT1-6His, or pEAQ-HT:16OMT2-6His, or pEAQ-HT: CRO_T104094-6His, or pEAQ-HT: FlavoOMT-6His or empty vector (EV). Black stars: proteins of interest. Red arrowhead: 16OMT1 homodimers. Green arrowheads: GFP protein.



Supplemental Figure S11 - MIA accumulation in plants transiently overexpressing an empty vector (EV), 16OMT1, 16OMT2 or T16H2. The 16-methoxytabersonine and vindoline concentration in *C. roseus* infiltrated leaves was measured 7 days after infiltration. Points correspond to individual samples, in which 2 leaves per plant were analyzed (3 plants were transformed for each condition). The individual samples were analyzed by ANOVA followed by the post hoc Tukey test. Same letters express no significant difference between the means of the conditions at 5% of significance. MIA composition is expressed in relative quantification (Peak area/leaf dry weight).



Supplemental Figure S12. Genomic locations of the *C. roseus* OMT candidates depicted with red arrows. **A)** 16OMT1 (CRO_T110596) was located on scaffold cro_v2_scaffold_16 within 14.8 kBp from CrT16H2 (CRO_T110598), and 22.6 kbp from CrT16H1 (CRO_T110599) **B)** 16OMT2 (CRO_T104105) and CRO_T104091 were both located 135 kbp apart on scaffold cro_v2_scaffold_11.

Supplemental Table 1. Primers used in this study.

	Name	Fw sequence	Rev sequence
Yeast expression primers	pESC_16OMT1	CTGAGAACTAGTATGGATGTTCAATCTGAGGAGTTCC	CTGAGAACTAGTAGGATAAACCTCAATGAGACTCCTTG
	pESC_CROT104105	CTGAGAACTAGTATGGATGTTCAATCGGATGAGTTCC	CTGAGAACTAGTTCAAGGATAAATCTCAATAAGACTCCTTGTG
	pESC_CROT104094	CTGAGAACTAGTATGGAAGTTCAATCAGTCGAGTTT	CTGAGAACTAGTTTAAGGATAAACCTCAATAATACTCCTATAATC
Recombinant proteins expression	pRSETA_16OMT1	CTGAGAGGATCCATGGATGTTCAATCTGAGGAGTTCC	CTGAGAGGATCCAGGATAAACCTCAATGAGACTCCTTG
	pRSETA_CROT104105	CTGAGAGGATCCATGGATGTTCAATCGGATGAGTTCC	CTGAGAGGATCCTCAAGGATAAATCTCAATAAGACTCCTTGTG
	pRSETA_CROT104094	CTGAGAGGATCCATGGATGTTCAATCGGATGAGTTCC	CTGAGAGGATCCTCAAGGATAAATCTCAATAAGACTCCTTGTG
	pRSETA_FlavoOMT	CTGAGAGGATCCATGGAATTGCAATCTTCTGAGATTC	CTGAGAGGATCCTCAAGGATATACTTCAATAGGACTTCTAAAATCT
Subcellular localization/ BiFC studies	locali_16OMT1	CTGAGAACTAGTATGGATGTTCAATCTGAGGAGTTCC	CTGAGAACTAGTAGGATAAACCTCAATGAGACTCCTTG
	locali_CROT104105	CTGAGAACTAGTATGGATGTTCAATCGGATGAGTTCC	CTGAGAACTAGTAGGATAAATCTCAATAAGACTCCTTGTGAA
	locali_CROT104094	CTGAGAACTAGTATGGAAGTTCAATCAGTCGAGTTT	CTGAGAACTAGTTCAAGGATAAATCTCAATAAGACTCCTTGTG
Gene expression analysis (qPCR)	qPCR_16OMT1	AATGGGCATTTCTCTTTAAGGA	CCGTAATACAAATTGGGTACAA
	qPCR_CROT104105	AGGGCAATTCCAGAAAAGGAAAA	CATAGATATCTGTGTTTTAACAGC

	qPCR_CROT104094	CCTTAATTATTATGATTAGATATATAATTTT	ATCTTGTGTGTGTCTCTAGACATCT
	qPCR2_16OMT1	TGTTTGAGGGCTTGGCTTCT	TGAGGAAAATTCTTGGCAATGG
	qPCR2_16OMT2	AGGAAAAGGGTGGAAAGGTGAT	CTGTGTTTTAACAGCATTGTCATCTTC
	qPCR_STR	CATAGCTCTGTGGGTATATTAGTGT	ACCTTCAAAAAGCCATAGAATTTGCA
	qPCR_SGD	CTTCGACAACTTCGAATGGAA	CTTCTTGACTAACTCAACTAGT
	qPCR_T16H2	GATCAACTCACAGTGGCAGTC	GACTTGAGGACTTGTGATTGGC
	qPCR_RPS9	TTACAAGTCCCTTCGGTGGT	TGCTTATTCTTCATCCTCTTCATC
	qPCR_HDS	ATTATGGCGACCGGAACAGTTCCA	GGCTACATTCCCAACCATTACTGTGCGTGT
Virus induced silencing (VIGS)	vigs_16OMT1	TCTAGACCGGGGGACCACAAACTTA	CTCGAGGTTCAGAACTATAACCCCAGAA
	vigs_16OMT2	TCTAGAGAAAATCGCCATGGATGTT	CTCGAGGAAATCAATTGGGAATCGC
	vigs_OMTs	CTGAGAGGATCCTTATAGTTCTGAACATATGGAACATG	CTGAGAGGATCCCTATACATTACCATGTCTATGATTATCAC
Transient agro-mediated overexpression	pEAQ_16OMT1	CTGAGACCCGGGATGGATGTTCAATCTGAGGAGTTCC	CTGAGACCCGGGAGGATAAACCTCAATGAGACTCCTTG
	pEAQ_16OMT2	CTGAGACCCGGGATGGATGTTCAATCGGATGAGTTC	CTGAGACCCGGGAGGATAAATCTCAATAAGACTCCTTGTGAA
	pEAQ_CROT104094	CTGAGAGGATCCATGGAAGTTCAATCAGTCGAGTTT	CTGAGAGGATCCTTAAGGATAAACCTCAATAATACTCCTATAATC

	pEAQ_FlavoOMT	CTGAGAACCGGTATGGAATTGCAATCTTCTGAGATTC	CTGAGAACCGGTAGGATATACTTCAATAGGACTTCTAAAATCTATC
	pEAQ_T16H2	CTGAGAACCGGTATGGAGTTGTATTATTTTTCCACC	CTGAGAACCGGTATATTTACCTTTGAGAGAAGAAGCA
Miscellaneous	pEAQ_NBCROT104094 (Nicotiana benthamiana expression)	CTGAGAACCGGTATGGATGTTCAATCGGATGAGTTCC	CTGAGAACCGGTAGGATAAATCTCAATAAGACTCCTTGTG
	QCROT104094_fusion (internal primers)	AAATTAGATGGAGAAACAGAATCTTATTATTCTCTAACCCCTTCAAGC	ATAAGATTCTGTTTCTCCATCTAATTTGTCTTGAGAAAAGAAGCCTG

Supplemental Table 2. Galactose-induced yeast strains used to functional validation of the candidate isoforms

Yeast Strain	Plasmids content
T16H2	pYEDP60_T16H2
T16H2 + OMT1	pYEDP60_T16H2 + pESCLEU_16OMT1
T16H2 + OMT2	pYEDP60_T16H2 + pESCHIS_16OMT2
T16H2 + CRO_T104094	pYEDP60_T16H2 + pESCLEU_ CRO_T104094
T16H2 + QCRO_T104094	pYEDP60_T16H2 + pESCLEU_Q CRO_T104094

Article 4 : Identifying genes involved in alkaloid biosynthesis in *Vinca minor* through transcriptomics and gene co-expression analysis

In the article 4, I was responsible for assisting the biochemical characterization of the 16OMT orthologue (Vm16OMT) that was identified by transcriptome studies developed by other authors.

RESUME EN FRANÇAIS DE L'ARTICLE 4

La petite pervenche *Vinca minor* accumule de nombreux alcaloïdes indoliques monoterpéniques (AIM) dont le vasodilatateur vincamine. Alors que la voie de biosynthèse des AIM a été largement élucidé chez d'autres Apocynaceae comme *Catharanthus roseus*, elle reste largement méconnue chez *V. minor*, notamment pour les réactions conduisant à des AIM spécifiques à cette plante. Par conséquent, nous avons généré un transcriptome complet de *V. minor* élaboré à partir de huit échantillons comprenant les racines, les feuilles jeunes et âgées exposées faiblement ou fortement à la lumière. Ce transcriptome est plus complet que ceux publiés précédemment. Grâce à des recherches basées sur l'homologie de séquences avec les gènes de *C. roseus*, nous avons prédit les gènes candidats pour toutes les étapes de la voie des AIM communes aux deux espèces. Ce résultat est illustré par le clonage d'une isoforme tabersonine/vincadiformine16-O-méthyltransférase (Vm16OMT). La validation fonctionnelle de cette enzyme a révélé sa capacité à méthyler des dérivés 16-hydroxylés de la tabersonine, de la vincadiformine et de la lochnéricine avec un Km de 0,94 ± 0,06 µM pour la 16-hydroxytabersonine. De plus, en réalisant des fusions avec des protéines fluorescentes et des études d'interaction, nous avons montré que la Vm16OMT est cytosolique et forme des homodimères. Enfin, un réseau de co-expression de gènes a été réalisé afin d'identifier les gènes candidats des étapes de biosynthèse manquantes de *V. minor* en vue de guider l'élucidation de la voie des AIM chez cette plante.

biomolecules

MDPI

Article

Identifying Genes Involved in Alkaloid Biosynthesis in *Vinca minor* through Transcriptomics and Gene Co-Expression Analysis

Emily Amor Stander [1,†], Liuda Johana Sepúlveda [1,2,†], Thomas Dugé de Bernonville [1], Inês Carqueijeiro [1], Konstantinos Koudounas [1], Pamela Lemos Cruz [1], Sébastien Besseau [1], Arnaud Lanoue [1], Nicolas Papon [3], Nathalie Giglioli-Guivarc'h [1], Ron Dirks [4], Sarah Ellen O'Connor [5], Lucia Atehortùa [2], Audrey Oudin [1,*] and Vincent Courdavault [1,*]

1 EA2106 "Biomolécules et Biotechnologies Végétales", Université de Tours, 37200 Tours, France; emily.stander@univ-tours.fr (E.A.S.); liudayohana@gmail.com (L.J.S.); thomas.duge@univ-tours.fr (T.D.d.B.); ines.carqueijeiro@univ-tours.fr (I.C.); koudounas@univ-tours.fr (K.K.); pamela.lemos@univ-tours.fr (P.L.C.); sebastien.besseau@univ-tours.fr (S.B.); arnaud.lanoue@univ-tours.fr (A.L.); nathalie.guivarch@univ-tours.fr (N.G.-G.)

2 Laboratorio de Biotecnología, Sede de Investigación Universitaria, Universidad de Antioquia, Antioquia Medellin 050021, Colombia; latehor@gmail.com

3 Host-Pathogen Interaction Study Group (GEIHP, EA 3142), UNIV Angers, UNIV Brest, 49933 Angers, France; nicolas.papon@univ-angers.fr

4 Future Genomics Technologies, 2333 BE Leiden, The Netherlands; dirks@futuregenomics.tech

5 Department of Natural Product Biosynthesis, Max Planck Institute for Chemical Ecology, 07745 Jena, Germany; oconnor@ice.mpg.de

* Correspondence: audrey.oudin@univ-tours.fr (A.O.); vincent.courdavault@univ-tours.fr (V.C.)

† These authors contributed equally to this paper.

Received: 26 October 2020; Accepted: 21 November 2020; Published: 24 November 2020



Abstract: The lesser periwinkle *Vinca minor* accumulates numerous monoterpene indole alkaloids (MIAs) including the vasodilator vincamine. While the biosynthetic pathway of MIAs has been largely elucidated in other Apocynaceae such as *Catharanthus roseus*, the counterpart in *V. minor* remains mostly unknown, especially for reactions leading to MIAs specific to this plant. As a consequence, we generated a comprehensive *V. minor* transcriptome elaborated from eight distinct samples including roots, old and young leaves exposed to low or high light exposure conditions. This optimized resource exhibits an improved completeness compared to already published ones. Through homology-based searches using *C. roseus* genes as bait, we predicted candidate genes for all common steps of the MIA pathway as illustrated by the cloning of a tabersonine/vincadifformine 16-*O*-methyltransferase (Vm16OMT) isoform. The functional validation of this enzyme revealed its capacity of methylating 16-hydroxylated derivatives of tabersonine, vincadifformine and lochnericine with a Km 0.94 ± 0.06 µM for 16-hydroxytabersonine. Furthermore, by combining expression of fusions with yellow fluorescent proteins and interaction assays, we established that Vm16OMT is located in the cytosol and forms homodimers. Finally, a gene co-expression network was performed to identify candidate genes of the missing *V. minor* biosynthetic steps to guide MIA pathway elucidation.

Keywords: alkaloids; *Vinca minor*; RNA-seq; transcriptomics; differential gene expression; *O*-methyltransferase

1. Introduction

Monoterpene indole alkaloids (MIAs) are a class of specialized plant metabolites found in Gentianale order families; Apocynaceae, Rubiaceae and Loganiaceae. MIAs have important roles in plant defense [1–3] and constitute a large chemical diversity of pharmacophores with a variety of potent biological activities. These activities make MIAs attractive pharmaceutical drugs with high commercial value, such as the anticancer drugs vinblastine and vincristine from *Catharanthus roseus*, the antiarrhythmic agent ajmaline from *Rauvolfia serpentina* and the vasodilator vincamine, from *Vinca minor*, used to treat cerebrovascular disorders. MIA producing plants vary a lot in biomass production and alkaloid content, prompting research efforts towards biosynthetic pathway elucidation. This knowledge has the potential to stimulate alternative engineering approaches, including synthetic biology and metabolic engineering, for the upscaled commercial production of these valuable natural products [4–6].

V. minor, also known as the lesser periwinkle, is an evergreen herbaceous shrub belonging to the Apocynaceae family. In addition to the above mentioned pharmaceutically important MIA vincamine, this plant produces more than 50 indole alkaloids, including t less known interesting molecules such as lochnericine, akuammicine, vincadifformine, minovincinine [7,8] and picrinine [9]. To date, publicly available *V. minor* genomic resources are restricted to the 454 based- (V_minor_v1) and illumina based transcriptomes (V_minor_v3) from the PhytoMetaSyn database (https://bioinformatics.tugraz.at/phytometasyn/). Furthermore, knowledge on MIA biosynthetic pathways in *V. minor* is lacking and mainly results from homology-based identification of genes with its relative *C. roseus* sharing high genetic and metabolic similarities [8,10]. Recently, Verma et al. (2020) [10] identified 12 putative MIA biosynthetic genes in *V. minor* originating from either the *Rauwolfia* raucaffricine-type alkaloid pathway, or the vindoline pathway in *C. roseus*. Furthermore, a *V. minor* counterpart of the *C. roseus* TPT2 transporter, involved in the secretion of MIAs to the leaf surface, has also been characterized [8] together with a picrinine *N*-methyltransferase involved in ervincine biosynthesis [11]. At the metabolic level, while the intracellular distribution of MIAs is still a matter of debate in leaves of *V. minor* [8,12], experimental evidence tends to confirm that MIA metabolism is higher in younger tissues and notably young leaves [13–15]. In addition, vincamine has been described as the main alkaloid found in the leaves of *V. minor* (0.13%) whereas 1,2-dehydroaspidospermidine is the major alkaloid detected in the roots (0.28% [16]). Finally, some alkaloids are also specific to *V. minor* leaves, including vincarubine, picrinine, eburnamine, eburnamonine, minovincinine, minovine, vincadifformine and strictamine [7,8,12,16].

Far from the poorly described biosynthetic routes of *V. minor*, a combination of pioneering works on *C. roseus* has culminated in the near-complete elucidation of MIA biosynthesis in the leaves of this "model" non-model plant [17–21]. By contrast, the MIA biosynthetic pathway in the roots of *C. roseus* still remains elusive to a certain extent [22–25]. Regardless of the MIA accumulating organs, almost all MIAs originate from a central unique skeleton, strictosidine, which is the product of a condensation reaction between tryptamine (the indole ring donor) and secologanin (the terpenoid donor) by strictosidine synthase (STR; Figure 1) [26,27].

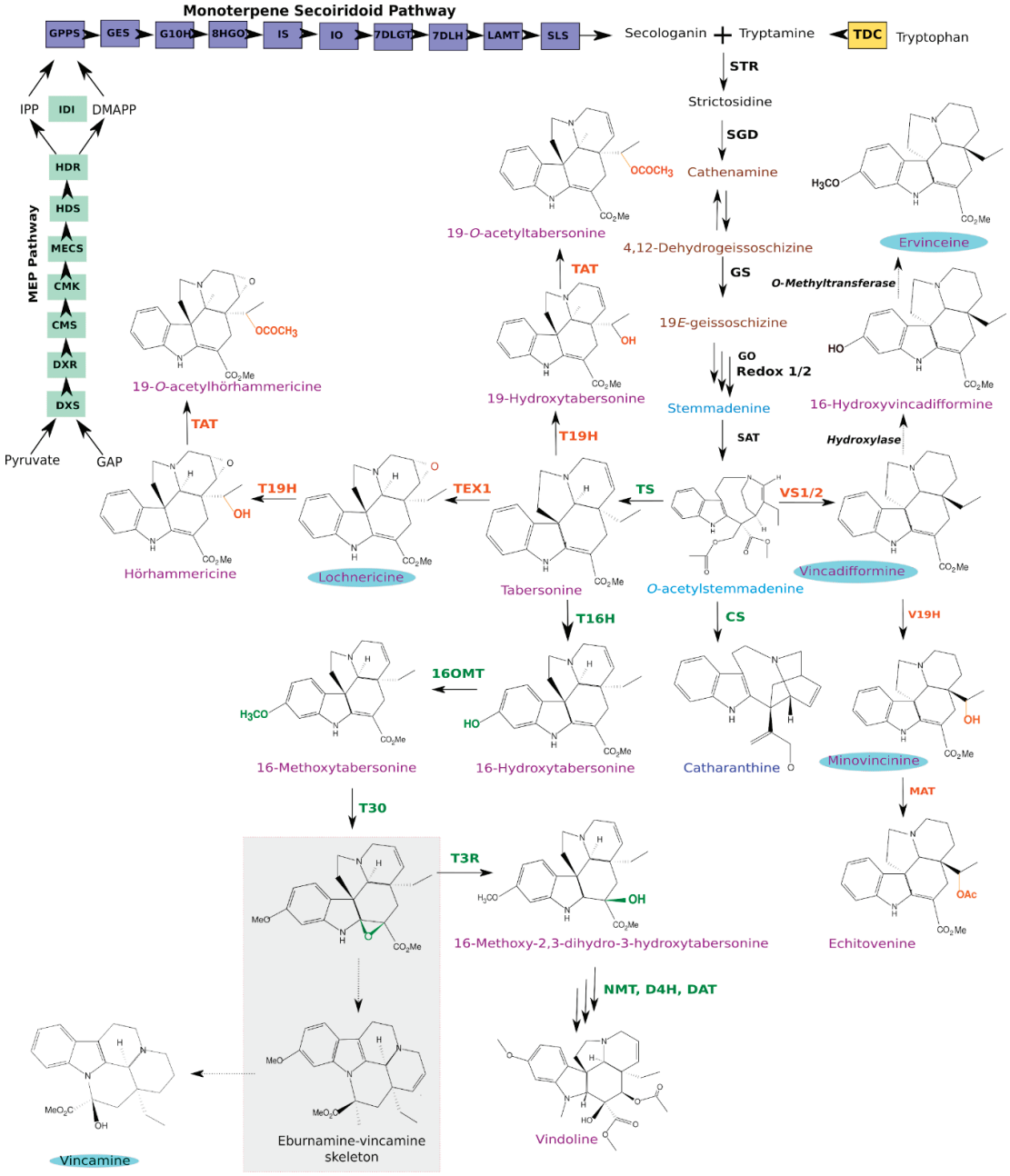


Figure 1. The elucidated MIA biosynthetic pathway in *Catharanthus roseus*. Enzymes catalyzing pathways in the aerial parts of the plant are depicted in green, and enzymes in the roots are depicted in orange. MIA names highlighted in light blue circles have also been identified in *Vinca minor*. Depicted in the grey box is the rearrangement of the T30 epoxide product, forming an eburnamine-vincamine like skeleton, in cases where T3R is silenced as described in [28]. The different MIA scaffold classes are depicted by MIA names written in dark blue (iboga-type), light blue (strychnos-type), brown (corynanthe) and purple (aspidosperma). Uncharacterized reactions are depicted with dashed arrows. MEP, methyl-D-erythritol phosphate; GAP, glyceraldehyde 3-phosphate; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; CMS, 2-C-methyl-D-erythritol 4-phosphate synthase; CMK, 4-(cytidine 5'diphospho)-2-C-methyl-D-erythritol kinase; MECS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, (E)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR,

(E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; IDI, isopentenyl diphosphate isomerase; GPPS, geranyl diphosphate synthase; GES, geraniol synthase; G10H, geraniol 10-hydroxylase; 10HGO, 10-hydroxygeraniol oxidoreductase; IO, iridoid oxidase; IS, iridoid synthase; 7DLGT, 7-deoxyloganetic acid glucosyltransferase; 7DLH, 7-deoxyloganic acid 7-hydroxylase; LAMT, loganic acid O-methyltransferase; SLS, secologanin synthase; TDC, tryptophan decarboxylase; STR, strictosidine synthase; SGD, strictosidine β-glucosidase; GS, geissoschizine synthase; GO, geissoschizine oxidase; SAT, Stemmadenine O-acetyltransferase; CS, catharanthine synthase; VS1/2, vincadifformine synthase 1/2; V19H, vincadifformine 19-hydroxylase; MAT, minovincine 19-O-acetyltransferase; TS, tabersonine synthase; TEX, tabersonine 6,7-epoxidase; T19H, tabersonine 19-hydroxylase; TAT, tabersonine derivative 19-O-acetyltransferase; T16H, tabersonine 16-hydroxylase; 16OMT, 16-hydroxytabersonine O-methyltransferase; T3O, tabersonine 3-oxidase; T3R, tabersonine 3-reductase; NMT, 16-methoxy-2,3-dihydrotabersonine N-methyltransferase; D4H, desacetoxyvindoline 4-hydroxylase; DAT, deacetylvindoline 4-O-acetyltransferase.

While tryptamine results from a single-step decarboxylation of tryptophan by tryptophan decarboxylase (TDC) [29,30], the synthesis of secologanin successively involves the methyl-erythritol phosphate (MEP) pathway and the monoterpene secoiridoid pathway producing secologanin. The MEP pathway successively involves 1-deoxy-D-xylulose 5-phosphate (DXP) synthase (DXS), DXP reductoisomerase (DXR), 4-(cytidine 5'diphospho)-2-C-methyl-D-erythritol synthase (CDP-ME) synthase (CMS), CDP-ME kinase (CMK), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECS), (E)-4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP) synthase (HDS) and HMBPP reductase (HDR). In association with IDI isomerase, it produces a controlled ratio of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) that are further condensed by geranyl diphosphate (GPP) synthase (GPPS) to form GPP [31–35]. The monoterpene secoiridoid pathway then converts GPP through the concerted action of geraniol synthase (GES) [36], geraniol 10-hydroxylase (G10H) [37], 10-hydroxygeraniol oxidoreductase (10HGO) [38], iridoid synthase (IS) [39], iridoid oxidase (IO) [40], 7-deoxyloganetic acid glucosyltransferase (7DLGT) [41], 7-deoxyloganic acid (7DLH) [42], loganic acid-O-methyltransferase (LAMT) [14] and secologanin synthase (SLS) [19,43]. The resulting strictosidine is then deglycosylated by strictosidine β-D-glucosidase (SGD) to form reactive intermediates, including 4,21-dehydrogeissoschizine and cathenamine, which initiate distinct MIA biosynthetic branches [44–47]. The gateway step in iboga- and aspidosperma-type MIA biosynthesis involves geissoschizine synthase (GS) that converts 4,21-dehydrogeissoschizine to 19*E*-geissoschizine [48]. This is followed by five additional enzymes that catalyze the biosynthesis of the common intermediate, *O*-acetylstemmadenine, which lies at the branching point between three different pathways [21,49]. In both above and underground parts of the plant, catharanthine synthase (CS) catalyzes the formation of the iboga MIA, catharanthine that is one of the structural monomers that joins with the aspidosperma MIA, vindoline to form precursors of the important anticancer drugs vinblastine and vincristine. Additionally, tabersonine synthase (TS) catalyzes the conversion of *O*-acetylstemmadenine towards the vindoline pathway via the important intermediate tabersonine [49]. Finally, in the roots of *C. roseus*, vincadifformine synthase 1 or 2 (VS1/2) catalyzes the formation of vincadifformine, which is converted into minovincinine by vincadifformine 19-hydroxylase (V19H) that in turn undergoes acetylation by minovincinine 19-hydroxy-O-acetyltransferase (MAT) to form echitovenine [22,25].

The metabolism of tabersonine, and to a lesser extent vincadifformine, is of primary importance in many Apocynaceae, notably in *C. roseus* and *V. minor*. For instance, tabersonine is converted into vindoline in leaves of *C. roseus* through a seven-step biosynthetic route [50]. These decorations start with an hydroxylation at the C16 position catalyzed by tabersonine-16-hydroxylase (T16H) [51–53], followed by a methylation at the same position by tabersonine 16-*O*-methyltransferase (16OMT), forming 16-methoxytabersonine [15,54]. A subsequent epoxidation of 16-methoxytabersonine at the C2/C3 position by 16-methoxytabersonine 3-oxygenase (T3O), results in an unstable epoxide [28], which in turn is converted to 16-methoxy-2,3-dihydro-3-hydroxy-tabersonine by tabersonine-3-reductase (T3R) [50].

This compound is then subjected to *N*-methylation, hydroxylation and acetylation respectively catalyzed by 16-methoxy-2,3-dihydro-3-hydroxytabersonine *N*-methyltransferase (NMT) [55], desacetoxyvindoline-4-hydroxylase (D4H) [56] and deacetylvindoline-4-*O*-acetyltransferase (DAT) [57] to form vindoline. Light was reported to be an important stimulator of the vindoline biosynthetic pathway [54,58] by activating the last two steps of the pathway involving D4H and DAT [56,58,59].

By contrast, in *C. roseus* roots, tabersonine is epoxidized at the C6, C7 position by tabersonine 6,7-epoxidase 1 (TEX1) to yield lochnericine [23], which is subsequently hydroxylated by tabersonine 19-hydroxylase (T19H) [60] and acetylated by tabersonine-derivatives acetyltransferase (TAT) [24] to form 19-*O*-acetylhörhammericine. Interestingly, a mirroring biosynthetic route allows converting vincadifformine into echitovenine through the concerted action of vincadifformine 19-hydroxylase (V19H) [25] and minovincinine acetyltransferase (MAT) [22]. While the presence of tabersonine and derivatives in *V. minor* has been reported in a few works only [8,61], vincadifformine and vincamine are accumulated at high levels [12]. Importantly, epoxide formation on a vincadifformine-like skeleton has been proposed to be part of the vincamine synthesis process leading to postulate that vincadifformine is a precursor of vincamine [62]. While orthologues of *C. roseus* VS1/2 can be easily identified, no enzyme involved in the vincadifformine/vincamine conversion has been identified to date. However, during T3O characterization, Kellner et al. (2015) [28] established that in the absence of T3R, the rearrangement of the epoxide product resulted in the accumulation of an eburnamine-vincamine skeleton product, which *C. roseus* does not naturally accumulate [63]. This thus suggested that a T3O like enzyme from *V. minor* could be involved in vincamine biosynthesis. Above all, it further highlights the necessity to pursue identification of MIA biosynthetic genes in *V. minor*.

While the elucidation of alkaloid biosynthetic routes has always been a laborious task, emerging omics-based strategies now facilitate the identification of genes from these complex pathways, notably through transcriptomics and gene co-expression analyses combined to functional genomic approaches [6,64]. Over the last five years, the efficiency of such a strategy has been clearly established by elucidating many missing steps of the MIA biosynthetic pathway in *C. roseus*. In this respect, the value of a high-quality reference transcriptome in discovering novel isoforms was also demonstrated, by identifying a second secologanin synthase (SLS2) catalyzing the last step of the synthesis of the monoterpene precursor of MIAs [19]. Therefore, the identification of missing MIA enzymes and isoforms appears to be crucial to elucidate the complexity of MIA metabolism as illustrated for cytochrome P450 reductases whose isoforms are associated to primary or specialized metabolism [65], the organ-specific distribution of T16H1 and T16H2 [53] and the enantiospecificity of T19H and V19H [25]. Furthermore, identification of novel isoforms across species is an ongoing endeavor that may allow for the discovery of more efficient enzymes to be applied in synthetic biology approaches for increased biosynthetic rates in newly developed cell factories [66,67].

In our continuing effort to elucidate and engineer MIA biosynthesis, we present here a comprehensive consensus transcriptome resource for *V. minor* comprising eight different tissue types and experimental conditions. This new *V. minor* reference transcriptome displays improved BUSCO scores to the previously available 454 based (V_minor_v1) and illumina based assemblies (V_minor_v3) from the PhytoMetaSyn database (https://bioinformatics.tugraz.at/phytometasyn/). This improved resource was thus used to rationally predict candidate genes for all known steps of the *V. minor* MIA pathway as illustrated by the functional validation of a tabersonine/vincadifformine 16-*O*-methyltransferase (16OMT) isoform. Finally, a gene co-expression analysis was performed to predict new MIA biosynthetic candidate genes to deepen our knowledge on the MIA metabolism in *V. minor*.

2. Materials and Methods

2.1. Chemicals

Vincadifformine was prepared by the hydrogenation of tabersonine as described in [68]. Lochnericine, 19-hydroxytabersonine and 16-hydroxytabersonine were produced by the 6,7-epoxidation of tabersonine, the 19-hydroxylation or 16-hydroxylation of tabersonine by yeast expressing TEX1, T19H or T16H2 from *C. roseus*, respectively [23,53]. S-adenosyl-L-methionine (SAM) was purchased from Sigma-Aldrich (St-Quentin-Fallavier, France).

2.2. Plant Growth Conditions and Sample Collection

Wild growing *V. minor* plants were collected from the center of France (47°21′11.4″ N 0°42′08.2″ E) and transferred to greenhouse (24 °C; 16 h light/8 h dark cycle). Stems were all removed to allow new shoots to develop during two months. In these conditions, two batches of ten plants were exposed to high (3300 lux) or low (250 lux) illumination conditions to generate the light and shadow conditions. Young leaves (YL) and old leaves (OL) stand for the first and fourth pair of leaves of these plants, respectively. Young leaves were collected from plants exposed to the two light exposure conditions to generate the light and shadow YL and YS, respectively. Flower buds (FB) were collected on plants exposed to high illumination conditions. Adventitious roots (AR) were induced as described in [24] using two-month old stems and were generated over one month. Two YL, YS and OL were collected per plant and each pair was considered as one biological replicate. Ten AR were gathered to generate one biological replicate (5 in total).

2.3. RNA-Sequencing

RNA was extracted from two biological replicates for AR, YL, OL and YS, using TRIzol (Thermo Fisher Scientific, Illkirch-Graffenstaden, France) as previously described [3]. The RNA was subsequently used to prepare libraries with the Illumina TruSeq RNA Library Prep Kit v2. Libraries were sequenced in paired-end mode (2 × 100 bp) by Eurofins Genomics (Les Ulis, France) using the Illumina HiSeq2500 technology. Sequencing data was deposited as project accession number PRJEB40906. EdgeR R package was used to process RNA-seq data. Replicate homogeneity was checked using a multidimensional scaling analysis (Figure S1). It clearly indicated that biological replicates were consistent. The biological coefficient of variation was 0.4 (common dispersion = 0.168531), in the range of typical variation for human data. Spearman rho between similar tissue types was >0.9 (Figure S2). We therefore assumed that in our case, only two biological replicates were sufficient for the purpose of this study.

2.4. Generation of a Consensus Transcriptome

Reads were trimmed using FastP [69] version 0.20 with default settings and used to de novo assemble individual transcriptomes using Bridger V2014-12-01 [70]. In order to construct a high quality reference transcriptome suitable for gene expression analysis and biosynthesis gene discovery, all assemblies were merged. Sequencing reads were pseudo aligned onto the merged assembly and counted with Salmon v0.14.1 [71] with bias correction (-biasCorrect) in the variational Bayesian optimized (-vbo) mode to establish the abundance estimates as transcripts per million (TPM). In our compaction procedure we discarded weakly expressed transcripts (>TPM5) and applied a clustering algorithm MMSeqs2 [72] using the following parameters: -mode 1, -c 0.8. Assemblies were evaluated with BUSCO [73] version 3.0.2 compared to existing *V. minor* transcriptomes (Vinca_minor_v1 and Vinca_minor_v3) obtained from the Phytometasyn database (https://bioinformatics.tugraz.at/phytometasyn/). Annotation of the consensus transcriptome was achieved with the Trinotate pipeline [74] version 3.0.1 that uses Blastx and BlastP of TransDecoder [75] V5.3.0 predicted ORFs against Uniprot, and hmmscan [76] against the PFAM database (https://pfam.xfam.org/).

2.5. Functional Analysis of the Consensus Transcriptome

Differentially expressed genes (DEGs) were identified by fitting a generalized linear model to each transcript using the edgeR R package [77]. Genes were considered to be differentially expressed if the *p*-value of the exact test on the binomial fit was below 0.01. In this analysis, genes with a log fold change above 2 were considered to be upregulated and genes with a log fold change below -2 were considered to be downregulated. Overlap in DEG sets between different comparisons were visualized by the UpSetR [78] package v1.4.0. Uniprot description keywords were retrieved from the transcriptome annotation and used in the enrichment tests by comparing effectives to a hypergeometric distribution with the phyper function in R [79].

MIA gene conservation was determined through blastn [80] analysis of the *V. minor* consensus transcriptome against known MIA genes from *C. roseus*. Transcripts with the highest bit scores were selected as representatives for each MIA gene equivalent. Graphics were generated with the ggplot2 R package [81].

2.6. Heterologous Expression of Tabersonine 16-O-Methyltransferase Candidates in Yeast

Dedicated primers (Table S3) were used to amplify the full coding sequences of T16H, YS1_comp1836_seq0, YS2_comp61_seq0 (Vm16OMT) and Cr16OMT. Primers were designed to include appropriate restriction sites at the extremities of each ORF for subsequent ligation into pESC-His plasmids. Each OMT expressing recombinant vector was cotransformed with the T16H2 expression vector into *S. cerevisiae* strain WAT11 as described in [53]. Transformed yeast strains were grown until the stationary phase in 4 mL of CSM medium (0.67% (*w*/*v*) yeast nitrogen base, 2% (*w*/*v*) dextrose and 0.05% (*w*/*v*) dropout mix without histidine) and harvested by centrifugation. Induction of protein expression was achieved by further culturing the harvested yeasts for 6 h in 10 mL of YPGal medium (2% bacto peptone, 1% yeast extract and 2% Gal) as described by [65].

2.7. Recombinant Vm16OMT Production in E. coli, Substrate Specificity Assays and Kinetic Parameters Determination

The Vm16OMT coding sequence was amplified with primers vm16OMTlike-for/vm16OMTlike-rev and cloned into the *BamHI* site of the pRSET-A vector (Thermo Fisher Scientific, Illkirch-Graffenstaden, France) before transformation of *E. coli* BL21(DE3) cells. Transformed cells were cultured until exponential growth (Abs595 nm = 0.5) and protein expression was induced by adding 1 mM IPTG for 4 h at 28 °C. The resulting recombinant protein was purified with a Co2+ matrix according to the manufacturer's instructions (Talon Resin metal ion, Clontech), concentrated and dialyzed against a phosphate buffer (50 mM pH 7.5) with Amicon®Ultra-4 (30 kDa cut-off, Millipore, Molsheim, France). The purified protein was aliquoted and 10% glycerol added before being flash frozen in liquid nitrogen. Protein concentration and purity were determined by the Bradford assay and SDS-PAGE, respectively.

Enzymatic activity assays were conducted at 30 °C by coincubating 1 µg of each recombinant protein in a final volume of 50 µL potassium phosphate buffer (50 mM pH 7 supplemented with 25 mM of ascorbic acid) with 1 mM S-adenosyl-L-methionine (SAM) and 1 µM of each respective substrate: tabersonine, lochnericine, vincadifformine and 19-hydroxytabersonine. The reactions were stopped with 1 (*v*/*v*) of 100% methanol, centrifuged at 13,000× *g* for 10 min and analyzed by LC–MS. The kinetic parameters were determined under the same conditions by varying substrate concentrations. For 16-hydroxytabersonine kinetics, the alkaloid concentration varied from 0.2 to 100 µM in the presence of a done quantity of SAM. For the SAM kinetics, the concentration of 16-hydroxytabersonine was set and the concentration of SAM varied between 1 and 200 µM. All reactions were performed in four biological replicates. The kinetic parameters were calculated from the enzyme initial activity measured using a non-linear regression following the Michaelis–Menten equation (SigmaPlot12).

2.8. UPLC-MS Analyses

UPLC-MS analyses were performed according to [65]. The selected ion monitoring mode was used to collect data for the following compounds: tabersonine (*m*/*z* 337), retention time (RT) = 12.36 min (RT) = 11.3 min; 16-hydroxytabersonine (*m*/*z* 353), RT = 7.9 min; 16-methoxytabersonine (*m*/*z* 367), RT = 12.04 min; 19-hydroxytabersonine (*m*/*z* 353), RT = 6.8 min; vincadifformine (*m*/*z* 339), RT = 11.16 min; lochnericine (*m*/*z* 353), RT = 11.2 min; 16-methoxylochnericine (*m*/*z* 383), RT = 11.4; Ervinceine (*m*/*z* 369) and RT = 11.5 min in Figure 7. Ervinceine was quantified as vincadifformine equivalent.

2.9. Gene Expression Measurements Using Real-Time RT-PCR

Expression levels of Vm16OMT were determined by quantitative reverse transcription-PCR according to [65] using dedicated primers compiled in Table S3, ACTIN was chosen as endogenous reference gene, and reverse transcribed RNAs from different *V. minor* organs, including young leaves (YL), old leaves (OL), flower buds (FB) and roots (R). qPCR were performed in triplicate using three biological replicates.

2.10. Subcellular Localization Studies

The full length coding sequence of Vm16OMT was amplified using a dedicated primer pair (Table S3) designed to introduce *SpeI* restriction sites to both extremities of the amplicon. The PCR product was subsequently cloned into either *SpeI*- or compatible *NheI* restriction sites of the pSCA-cassette YFPi [82] producing Vm16OMT-YFP and YFP-Vm16OMT fusion proteins, respectively.

For the bimolecular fluorescence complementation (BiFC) assays, the Vm16OMT cDNA was cloned into pSPYCE (MR), pSPYNE173 and pSPYCE (M) plasmids, at the *SpeI* restriction sites of either 5′ or 3′ ends of the coding sequence of the *N*-terminal (YFP^N, residues 1–173) and *C*-terminal (YFP^C, residues 156–239) fragments of YFP [83]. This allows expressing the fusion proteins Vm16OMT-YFP^N, Vm16OMT-YFP^C and YFP^C-Vm16OMT that were used to study protein interactions.

The resulting plasmids were transiently cotransformed with a plasmid expressing the nucleocytoplasm CFP marker [84] into *C. roseus* cells by particle bombardment followed by fluorescence imaging as previously described by [85]. In brief, the *C. roseus* cells were bombarded with DNA-coated gold particles (1 μm) and a 1100-p.s.i. rupture disc at a stopping-screen-to-target distance of 6 cm, using the Bio-Rad PDS1000/He delivery system. After the bombardment, the cells were allowed to recover for 16 - 30 h prior to harvesting and observation. The resulting subcellular localization imaging was performed with an Olympus BX-51 epifluorescence microscope equipped with an Olympus DP-71 digital camera and a combination of YFP and CFP filters. The localization patterns of ± 50 observed cells were finally presented in this study.

2.11. Co-Expression Network

For each possible gene pair in the consensus transcriptome, the pairwise Pearson's correlation coefficient was calculated and subsequently ranked to determine the highest reciprocal ranks (HRR) as described previously [86,87]. The best co-expressed genes were captured for each known MIA pathway gene if their HRR < 100. The resulting associations were visualized using the igraph package in R [88] and communities of the closest co-expressed genes were determined using a fast greedy algorithm [89].

3. Results and Discussion

3.1. A Consensus and Refined Transcriptome Resource for V. minor

A total of eight libraries (Figure 2) were prepared from RNA extracted from four different tissue types including, mature leaves exposed to light (OL), young leaves exposed to light (YL), young leaves grown under low illumination conditions (YS) and adventitious roots (AR) developing from cut-stems immersed in water. Since the accumulation of many MIAs is stimulated by light, including vindoline

in *C. roseus* [54,58], *N*,*β*-d-glucopyranosyl vincosamide in *Psychotria leiocarpa* [90] and camptothecin in *Camptotheca acuminata* [91], we thus exposed *Vinca minor* plants to low and high light exposure conditions (namely shadow and light respectively), in order to explore whether a similar transcriptional regulation could be detected in *V. minor*. Adventitious roots were preferred to native roots as these are more prone to yield high-quality RNA and relevant sequencing results [24]. Illumina sequencing yielded between 24 and 33 million paired-end and processed reads that were used to de novo assemble each sample transcriptome individually using Bridger (Figure 2). All transcriptomes yielded complete BUSCO scores > 91% and missing rates < 5%. BUSCO scores for already published *V. minor* transcriptomes depended on the data type [92]. The 454 based assembly (Vinca_minor_v1, Figures 2 and 3) displayed a high proportion of missing BUSCOs (23.4%) while the Illumina based assembly (Vinca_minor_v3, Figure 2) had a complete BUSCO score of 84% as well as a relatively high fragmentation rate (11%) in contrast to our assemblies (<4% on average; Figure 2).

Next, we aimed at constructing a high-quality reference transcriptome suitable for gene expression analysis and discovery of MIA biosynthetic genes. We thus combined the individual transcriptomes resulting in a merged resource of 896,265 transcripts and a complete BUSCO score of 99.2% (all non compacted, Figure 2). However, this large number of transcripts is far from the closely related species *C. roseus* whose genome was predicted to contain 34,363 genes [93]. Therefore, we aimed to decrease the final merged assembly size by removing redundant sequences without losing biologically relevant transcripts. In our downstream compaction procedures, we discarded weakly expressed genes (TPM < 5) and applied a sequence clustering algorithm (MMSeqs2) [72].

Finally, this procedure allowed a reduction of the merged transcriptome (896,265 transcripts) to a final consensus transcriptome of 41,823 transcripts (all compacted) and an acceptable complete BUSCO score of 90.1% (Figures 2 and 3). This is the first description of a consensus transcriptome resource for *V. minor*. For comparison purposes, genome predictions for *Camptotheca* had a BUSCO score of 93.6% [94] and for *C. roseus* v2 93.3% [93]. Consequently, this *V. minor* consensus transcriptome was used for further analyses.

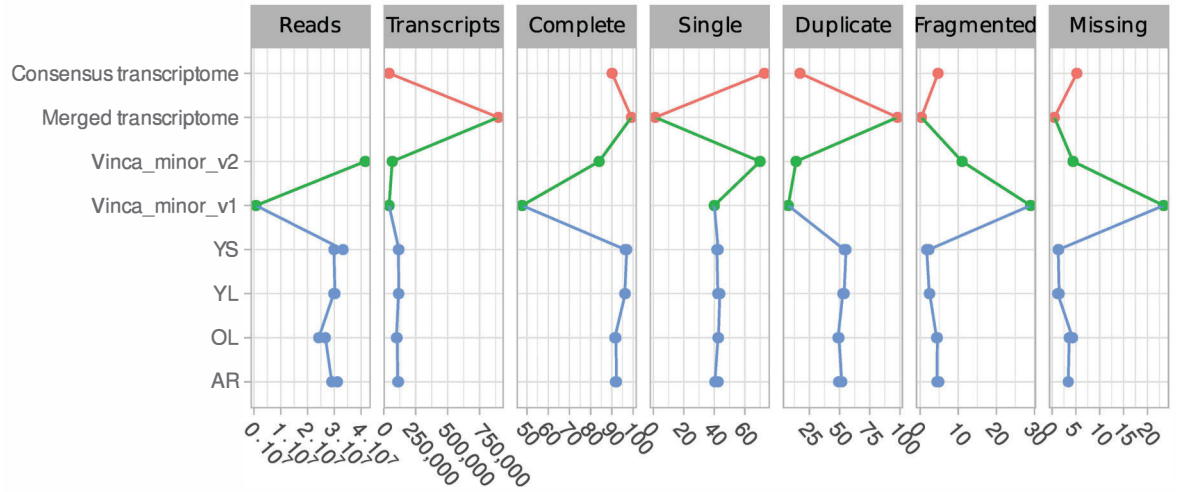


Figure 2. Comparison of *V. minor* transcriptomes in terms of number of generated reads, number of assembled transcripts and BUSCO evaluation scores (% complete, single, duplicated, fragmented and missing). Data sources are depicted as different colors with blue generated in this study, green obtained from previously published data and orange derived from our generated combined transcriptomes.

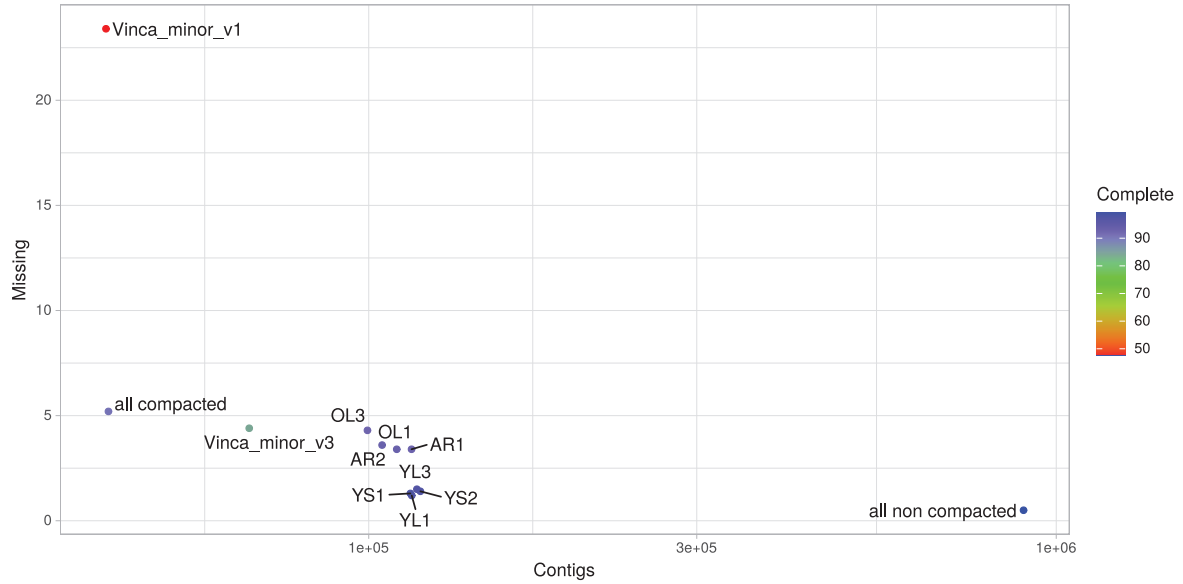


Figure 3. BUSCO evaluation of transcriptome assembly quality. The number of missing BUSCOs is plotted against the total number of contigs. Colors correspond to the percentages of completed BUSCOs. BUSCO analysis was conducted for each individual sample transcriptome for old leaves light (OL), young leaves light (YL), young leaves shadow (YS) and adventitious roots (AR), and their merged transcriptome without (all non compacted) and with compaction (all compacted). BUSCO metrics were compared with available *V. minor* transcriptome resources including the 454 based assembly (Vinca_minor_v1) and the Illumina based assembly (Vinca_minor_v3).

3.2. Functional Annotation of the V. minor Consensus Transcriptome and Differentially Expressed Gene Analysis

In Apocynaceae, MIA biosynthetic pathways exhibit complex distributions in different organs and cell-types, and was described in *C. roseus* [95] and predicted in *V. minor* based on MIA distribution [16]. A generalized linear model was thus constructed to detect differentially expressed genes (DEGs) in distinct organs by focusing on five comparisons: YL vs. AR, YS vs. AR, OL vs. AR, YL vs. YS and YL vs. OL (Figure 4). Logically, the most contrasting comparisons opposed aerial organs (YL, YS and OL) to adventitious roots (AR). We found 3799 transcripts significantly more expressed in aerial parts (YS + YL + OL vs. AR; Figure 4A), representing expected Gene Ontology (GO) biological processes, including photosynthesis and responses to light stimulus (Figure S3). In AR, the 2982 specific DEG (Figure 4B) were particularly found to be related to ion transport and homeostasis, oxidative stress, and lignin metabolism (Figure S2). Additionally, we found 2149 transcripts that were exclusively expressed in YL compared to OL (Figure 4A), implying specific metabolic activities (primary or specialized). In YL, a large portion (>100) of transcripts were identified that were related to cell division, mitotic cell cycle phase transition and mitotic cell cycle (Figure S5), revealing intense mitotic activity of YL in contrast to OL. These processes were also found to be increased in AR compared to OL (Figure S6), further emphasizing the active growth of these younger tissues, compared to OL.

The YL vs. YS sample comparison provided insights into the effects of light on young leaves of *V. minor*. Even though this comparison only contained 185 DEG, YL samples did contain genes associated with GO biological pathways (p-value < 0.05). Interestingly, light increased the expression of genes related to environmental stresses such as salt stress, heat stress, insects and reactive oxygen species (Figure S7). Furthermore, there was an enrichment of genes related to primary metabolisms, such as the starch catabolic process and lipid transport in YL compared to YS.

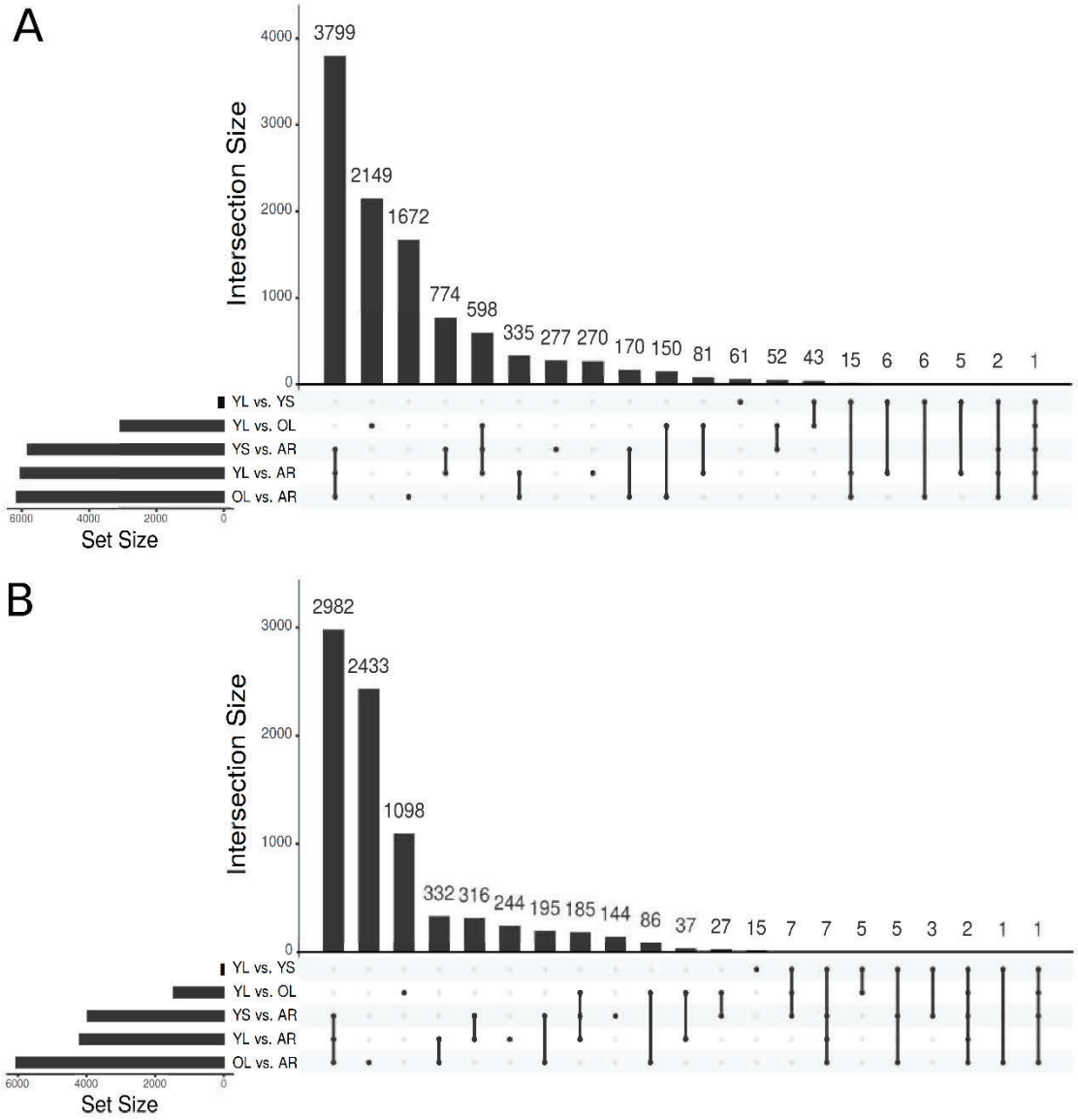


Figure 4. Differential gene expression analysis. UpSetR [78] plots highlighting the intersections resulting from the differentially expressed gene (DGE) analysis for the different comparisons as (**A**) log fold change > 2 indicating increased expression and (**B**) log fold change < −2 indicating decreased expression. Set sizes indicate the total number of transcripts differentially expressed for each comparison.

Finally, GO keyword enrichment analysis identified 32 transcripts related to alkaloid metabolism that were significantly enriched in YL compared to AR, and 19 alkaloid metabolism related transcripts significantly enriched (p-value < 0.05) in AR compared to YL (Figure S8). Furthermore, >20 transcripts with GO enrichment terms related to the alkaloid metabolism were identified in AR, compared to OL (Figure S9), and YL had more than double the number of transcripts enriched in alkaloid metabolism (20) compared to OL (Figure S10), indicating that alkaloid metabolism may be favored in these younger tissues. Such differences in alkaloid metabolism gene expression thus confirm previous results indicating that activities of MIA biosynthetic enzymes are higher in young tissues [13–15]. This contrasted expression also suggests that gene expression networks could result in new MIA gene identification.

3.3. Identification of Candidate Genes from the MIA Pathway through Homology-Based Predictions

The MIA pathway in *C. roseus* is now well described and shares high similarity with that of *V. minor* for all common steps up to vincadifformine or tabersonine, for instance. We thus conducted a homology-based identification of MIA genes by analyzing *C. roseus* gene conservation in our *V. minor* assemblies through blastn analysis followed by selecting the best Bit scoring transcripts for each hit. For all known MIA genes from *C. roseus* encoding biosynthetic enzymes, transporters and transcription factors, the best hits from our *V. minor* assembly were identified and listed in Table S1. Overall, MIA gene reconstruction was similar in each assembly but in some cases % identity was found to be better in the compacted *V. minor* resources (Figure 5A). Most single transcriptomes had at least one MIA hit with lower sequence identity than the transcript equivalent captured in the consensus assembly, as for example, TPT2 (YS1 samples), redox1 (OL1 samples), deoxyxylulose-5-phosphate synthase 1 (DXS1) from the MEP pathway (YL1 and AR1) and T3O (YL3, YS2, AR1 and AR2 samples). By combining different transcriptomes from different tissue types, we can anticipate that the best transcripts from each set will be captured. Interestingly, CMK in YL1 and YL3 had higher % amino acid sequence identity than the transcript captured in the consensus transcriptome (Figure 5A). However, upon further investigation it was found that although both transcripts were captured in the all non compacted transcriptome, they did not display the highest bit-scoring transcripts in this assembly and were thus discarded. Furthermore, these transcripts had a lower alignment length (<100 bp) to the best hit transcript (>1400 bp). This case study further demonstrated the potential of combining different transcriptomes in order to rationally increase the chances of identifying true gene homologs.

Upon further exploration of our consensus transcriptome, we found four groups of gene sequence conservation compared to *C. roseus* (Figure 5B,C). Each group corresponds to a specific subpart of the pathway, thus highlighting a modular conservation of this pathway in *V. minor*. The highest gene conservation (>85% in average) was observed for genes from early biosynthetic steps including MEP and monoterpene secoiridoid branches, such as GES, IS, IO, DL7GT, DL7H, LAMT up to SLS. Interestingly, genes related to Strychnos-type MIAs (GS and GO) together with redox1 and redox2 (Figure 1) were also part of this group, suggesting a low sequence evolution permissiveness towards the encoded enzymes. Surprisingly, while their functions are well-conserved, we noted that TDC, STR and SGD belong to the second group of sequence conservation (<80% in average), which could result in slightly different catalytic properties. This group also includes genes encoding alcohol dehydrogenases involved in tetrahydroalstonine and other heteroyohimbine MIA synthesis (Figure 5B). Such lower conservation of tetrahydroalstonine synthase (THAS) was quite expected since tetrahydroalstonine has not been detected to date in *V. minor*. However, these genes may encode enzymes involved in the synthesis of other non-described MIAs from *V. minor*. In addition, genes of the four last steps of tabersonine synthesis (SAT, PAS, DPAS and TS) also display lower identities, similarly to those observed for TDC, STR and SGD. This second group also includes vincadifformine synthase (VS1 and VS2) and orthologs of T16H1 and T16H2 that may catalyze the hydroxylation of tabersonine and/or vincadifformine. It thus gathers many orthologs of genes involved in the central steps of MIA biosynthesis, mainly synthesis and first decorations of strictosidine. The third group of gene conservation (around 75% identity) contains less orthologs, mostly related to putative 19-hydroxylation of tabersonine (V19H and T19H), vindoline (16OMT, T3R and NMT) and catharanthine (CS) synthases. Once again, these results were expected since no similar reactions have been described to date in *V. minor* but these genes likely correspond to candidates of interest for the synthesis of other and/or related MIAs. However, orthologs of T3O also belong to this group and may constitute potential candidates for the cyclization reaction of vincamine synthesis [28]. The fourth and last group of identity (below 70% of identity) mainly encompasses orthologues of lochnericine synthases (TEX1 and TEX2) that were unexpectedly quite distant from the *C. roseus* counterparts, and orthologues of the acyltransferases DAT, TAT and MAT. While no acetylated version of tabersonine has been described to date, these acyltransferases may be engaged in the synthesis of other acetylated MIA such as acetylated vincarubine [96]. Finally, for six of the candidate genes, we validated their expression levels quantified

by RNA-seq by quantitative PCR performed on three different biological samples (correlation 0.785, *p*-value = 5.701e^{-06}, Pearson CC), thus suggesting that gene co-expression analysis could be conducted to guide prediction/identification of MIA biosynthetic genes (Figure S11). In conclusion, analyzing our consensus transcriptome through homology-based searching resulted in an exhaustive identification of potential MIA biosynthetic genes in *V. minor*. Furthermore, it highlighted differential degrees of gene sequence conservation that might result from the intrinsic capacities of enzymes to evolve without losing initial activity and to acquire new or altered catalytic properties [97,98].

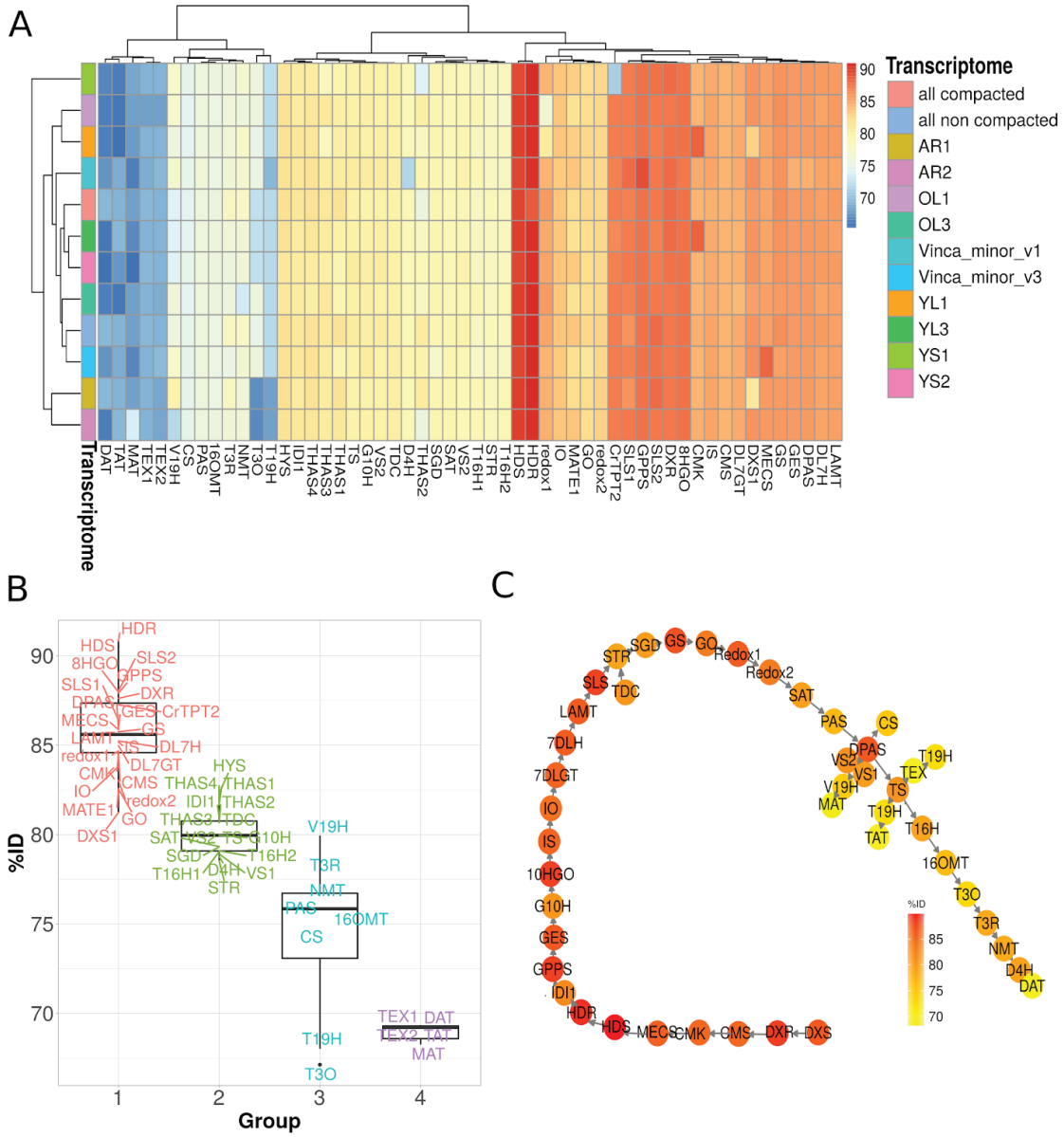


Figure 5. Comparison of *V. minor* transcripts to *C. roseus* reference sequences from the MIA pathway. (**A**) Percent of identity was calculated with blastn for each *V. minor* assembly against *C. roseus* MIA nucleotide sequences. (**B**) Average %ID per group of genes in the *V. minor* consensus assembly. (**C**) Comparison of *V. minor* consensus assembly homologs to the elucidated *C. roseus* vinblastine pathway.

3.4. Identification and Functional Validation of a V. minor Vincadifformine/Tabersonine 16-O-Methyltransferase

To determine the reliability of our annotated consensus transcriptome in identifying candidate MIA biosynthetic pathway genes, we performed the functional validation of one of the predicted candidates. We thus chose one from the third group of gene conservation to characterize a candidate with a

relatively low identity (75% in average) to strengthen the approach. More specifically, we focused on identifying an ortholog of 16OMT that catalyzes the methoxylation of 16-hydroxytabersonine in *C. roseus* leaves while it is notably proposed to methylate 16-hydroxyvincadifformine to produce ervinceine in *V. minor* (Figure 1). By mining our *V. minor* consensus transcriptome, we identified four tabersonine 16-*O*-methyltransferase-like candidates (Table S1) namely AR2_comp35_seq0, YS2_comp61_seq0, AR1_comp1185_seq0 and YS1_comp1836_seq0, by decreasing identity. All candidates display nucleotide sequence similarity (>70%) to the *C. roseus* 16OMT. The deduced protein sequences of all four candidates maintained *O*-methyltransferase domains (PF00891.15, Figure S12), as determined by hidden-Markov comparison against the PFAM database. Interestingly, the relative expression levels (as transcripts per million; TPM) of each candidate, obtained by mapping the reads from each tissue type to the consensus transcriptomes (Figure 6A), revealed that YS2_comp61_seq0 display the highest expression in all the organs tested with a maximum in the young aerial parts of the plant (YL and YS, >1300 TPM). This high level of expression was confirmed by quantitative RT-PCR providing additional evidence of high expression in YL (Figure 6B). By contrast, the three remaining candidates were barely detectable in old leaves and roots. Such differences in gene expression were of interest to prioritize candidate analysis given the specificity of MIA accumulation, notably ervinceine, in YL (Figure 6C).

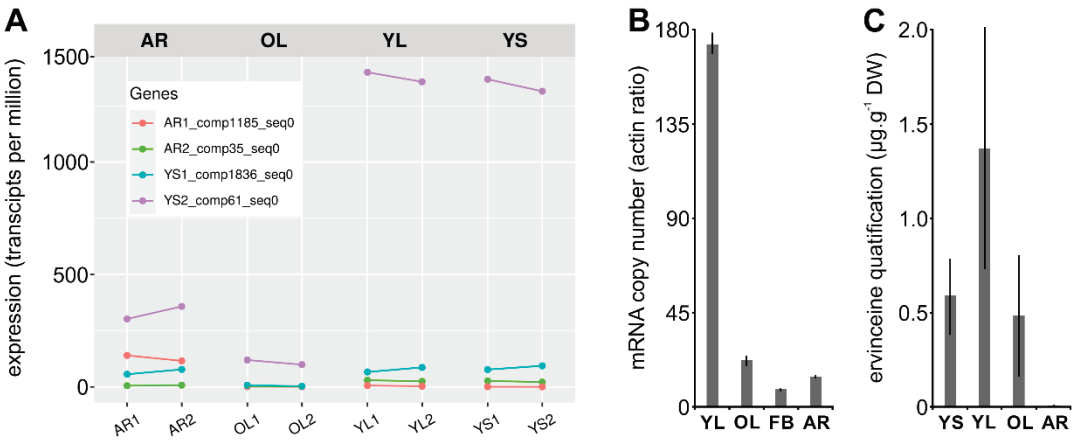


Figure 6. Relative transcript quantification of candidate tabersonine 16-*O*-methyltransferases and ervinceine amount measurement. (**A**) Relative expression was quantified as transcripts per million (TPM) after mapping RNA-seq reads from young leaves light (YL), young leaves shadow (YS), old leaves light (OL) and adventitious roots (AR) to the consensus transcriptome. (**B**) Relative YS2_comp61_seq0 quantification by reverse transcription quantitative PCR (qPCR) using RNA extracted from young leaves (YL), old leaves (OL), flower bud (FB) and roots (R). Data represents means ± SD of replicates performed on three distinct biological samples of each organ type. (**C**) Ervinceine quantification as µg/mg (mean ± SD, $n = 3$) in adventitious roots (AR), old leaves light (OL), young leaves light (YL) and young leaves shadow (YS) performed by UPLC–MS analysis.

Based on ervinceine accumulation and gene expression profiles (e.g., high expression in young leaves), we thus selected YS1_comp1836_seq0 and YS2_comp61_seq0 for further functional validation. Given 16OMT substrate promiscuity and its capacity to methylate distinct tabersonine derivatives [15,24], both candidates were first tested for tabersonine 16-*O*-methyltransferase activity by co-expression in yeast with T16H2. After cloning each coding sequence in pESC-HIS, yeasts were cotransformed with the corresponding plasmid together with pESC-LEU T16H2 [53]. As a control, yeast strains expressing either T16H2 or T16H2 and 16OMT from *C. roseus* were also included in the assay as previously described [24]. The resulting yeasts were then cultured, induced with galactose and fed tabersonine (125 µM) for 24 h, before analysis of the resulting products using ultra-performance liquid chromatography–mass spectrometry (UPLC–MS; Figure 7A). Comparisons of selected ions (*m*/*z* 337 for tabersonine, *m*/*z* 353 for 16-hydroxytabersonine and *m*/*z* 367 for 16-methoxytabersonine) to the

retention times of authentic standards allowed for the identification of the enzymatic reaction products. While yeasts expressing T16H2, T16H2 and 16OMT converted tabersonine into 16-hydroxytabersonine and 16-methoxytabersonine, respectively, we observed that only the YS2_comp61_seq0 candidate co-expressed with T16H2 consumed tabersonine and produced a compound with *m*/*z*, retention time and UV spectrum similar to 16-methoxytabersonine (Figure S13). This result thus confirms that YS2_comp61_seq0 displays tabersonine 16-*O*-methyltransferase activity and was subsequently named *V. minor* tabersonine 16-*O*-methyltransferase (Vm16OMT, GenBank accession no. MH010798). Above all, it confirmed the reliability of our annotated consensus transcriptome and predictions of MIA candidate genes in *V. minor*.

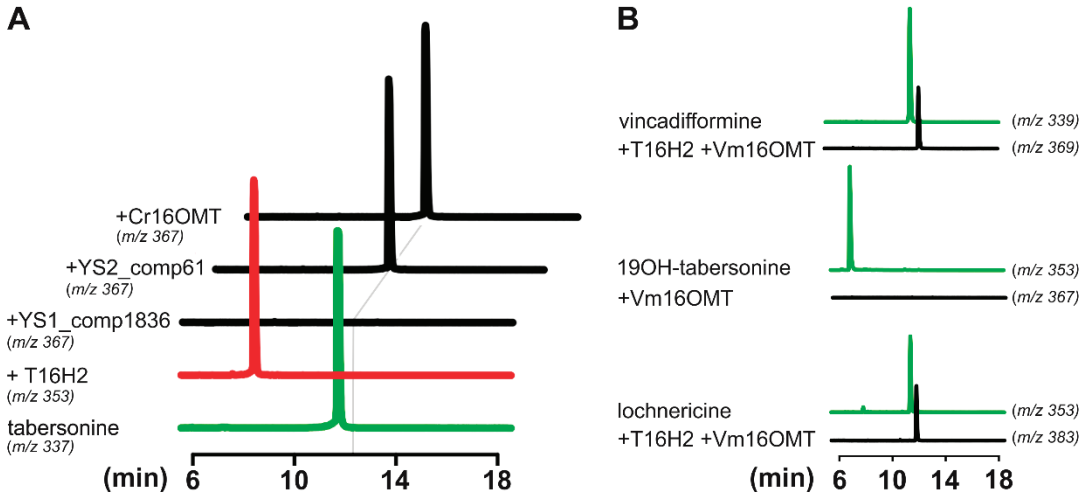


Figure 7. Screening selected candidates for tabersonine 16-*O*-methyltransferase activity. (**A**) Galactose induced yeast cell cultures co-expressing T16H2 (red line) with selected 16OMT-like candidates (YS1_comp1836_seq0 and YS2_comp61_seq0) and positive control Cr16OMT (black lines), incubated with tabersonine (green line, *m*/*z* 337). Selected ion monitoring and comparisons to authentic standards were used to detect the resulting hydroxylated (*m*/*z* 353) and methylated (*m*/*z* 367) tabersonine products by LC–MS analysis. The grey line highlights 16-methoxytabersonine retention time. (**B**) Substrate specificity of recombinant Vm16OMT in combination with recombinant T16H2 was determined by coincubation with 19OH-tabersonine, lochnericine and vincadifformine (green lines), forming methylated products that were detected by UPLC–MS (black lines). No methylated product was observed after incubation of recombinant Vm16OMT with 19-hydroxytabersonine.

3.5. Estimation of Vm16OMT Substrate Specificity

To determine the enzymatic substrate specificity of Vm16OMT in vitro, yeast strains expressing T16H2 and Vm16OMT (YS2_comp61_seq0) or Vm16OMT alone were fed with different substrates including vincadifformine/lochnericine and 19-hydroxytabersonine, respectively (Figure 7B). The resulting products were analyzed by UPLC–MS, which detected methylated products eluting at *m*/*z* 367 following tabersonine incubation, *m*/*z* 383 following lochnericine incubation and *m*/*z* 369 following vincadifformine incubation. As previously observed for *C. roseus* 16OMT, we concluded that Vm16OMT methylated the 16-hydroxylated tabersonine related compounds including 16-hydroxyvincadifformine and 16-hydroxylochnericine besides 16-hydroxytabersonine (Figure 7A,B). By contrast, no methylation of 19-hydroxytabersonine was observed. These results thus indicated that Vm16OMT can accept a variety of aspidosperma substrates but is regioselective towards the 16-position hydroxyl group as a methyl acceptor. The promiscuity of Vm16OMT towards a variety of 16-hydroxylated aspidosperma compounds further suggests that this enzyme may catalyze additional biosynthetic reactions in *V. minor*.

In order to gain insight into Vm16OMT enzymatic properties, its biochemical parameters were subsequently evaluated following heterologous expression in *Escherichia coli* and purification (Figure 8A; inset). The substrate specificity was determined by coincubating the recombinant Vm16OMT with

16-hydroxytabersonine as a methyl acceptor and S-adenosyl-L-methionine (SAM) as a methyl donor. Under these conditions, substrate saturation kinetics revealed that the recombinant Vm16OMT exhibited a high affinity for 16-hydroxytabersonine (Km 0.94 ± 0.06 µM; Vmax $3.35 \times 10^{-3} \pm 0.04 \times 10^{-3}$ µM·s^{-1}; kcat 0.0144 s^{-1}) and SAM (Km 46.2 ± 6.1 µM; Vmax $3.55 \times 10^{-3} \pm 0.16 \times 10^{-3}$ µM·s^{-1}; kcat 0.039 s^{-1}; Figure 8A,B), which remains in the same order of magnitude as Cr16OMT parameters [15]. This suggests that both OMTs share similar catalytic properties to ensure the synthesis of related MIAs in *C. roseus* and *V. minor*.

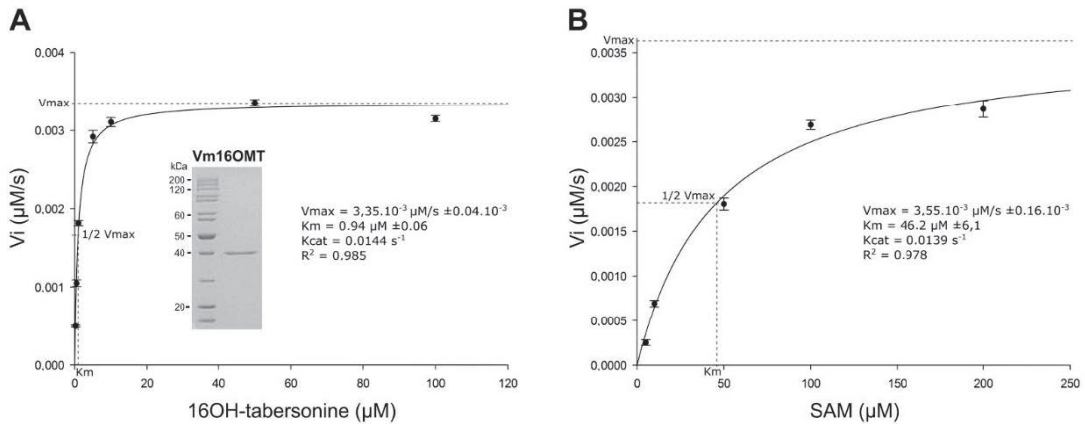


Figure 8. Determination of Vm16OMT kinetic parameters for 16-hydroxytabersonine (**A**) and SAM (**B**). Vm16OMT initial velocities in 16-methoxytabersonine synthesis were measured using fixed SAM concentration and from 0.2 to 100 µM 16-hydroxytabersonine (**A**) or fixed 16-hydroxytabersonine concentration and from 1 to 200 µM SAM (**B**). Nonlinear regressions based on the Michaelis–Menten equation were applied to estimate kinetic parameters. Purity of the recombinant Vm16OMT used was controlled on acrylamide gel stained with Coomassie blue (**A**) Error bars correspond to standard deviation of biological replicates (n = 4).

3.6. Subcellular Localization of Vm16OMT

While the cellular and subcellular organization of the MIA pathway has been widely characterized in *C. roseus* [19,95], only little is known about the corresponding architecture in *V. minor*. To gain insight into this organization and to complete the characterization of Vm16OMT, we studied the subcellular localization of Vm16OMT through YFP imaging analysis. The full-length Vm16OMT ORF was cloned at either 5′-end or 3′-end of YFP coding sequence in the pSCA-YFP plasmid to generate the Vm16OMT-YFP or YFP-Vm16OMT fusion proteins, respectively. Both constructs were transiently expressed in *C. roseus* cells in combination with a cyan fluorescent protein (CFP) nucleocytosolic marker (Figure 9A–H). For both orientations of fusion, a similar signal of fluorescence was observed, perfectly superimposed with that of the CFP marker. This suggests that Vm16OMT is mainly located in the cytosol and may undergo a passive diffusion into the nucleus due to its small size since no nuclear localization signal could be detected in Vm16OMT primary sequence. In addition, the quaternary structure of Vm16OMT was further analyzed through bimolecular fluorescence complementation (BiFC) assays. In this case, the Vm16OMT ORF was cloned at the 5′-end of the coding sequence of the split-YFP^N fragment (Vm16OMT-YFP^N) and either at the 5′- or at the 3′-end of the coding sequence of the split-YFP^C fragment to generate Vm16OMT-YFP^C and YFP^C-Vm16OMT, respectively. The two different combinations of constructs (Vm16OMT-YFP^N and Vm16OMT-YFP^C; Vm16OMT-YFP^N and YFP^C-Vm16OMT) were transiently expressed in *C. roseus* cells to analyze the reconstitution of the BiFC complex through the apparition resulting of a YFP fluorescent signal (Figure 9I–L). Such formation was observed for the Vm16OMT-YFP^N/Vm16OMT-YFP^C combination while no BiFC complexes were reformed with the Vm16OMT-YFP^N/YFP^C–Vm16OMT combination. These results thus suggest that Vm16OMT is engaged in homodimer or homomultimer formation through a head-to-head association. This was in agreement with the presence of a dimerization domain (PF08100.8, Figure S12) located at the

N-terminal end of plant *O*-methyltransferases [99]. Interestingly, we also noticed that Vm16OMT–BiFC complexes were localized exclusively in the cytosol and not in the nucleus as revealed by the lack of superimposition of the BiFC YFP signal with that of the nucleus CFP marker (Figure 9M–P). This nuclear exclusion probably results from the increase of Vm16OMT size due to homodimer formation (around 80 kDa) that may exceed nuclear pore exclusion size that is around 60 kDa [100]. Such restricted distribution to the cytosol has been already observed for *C. roseus* 16OMT and may favor captation of 16-hydroxytabersonine released by the endoplasmic reticulum-anchored T16H2 to increase MIA biosynthetic flux [53,84]. A similar hypothesis can be made for Vm16OMT cytosolic localization.

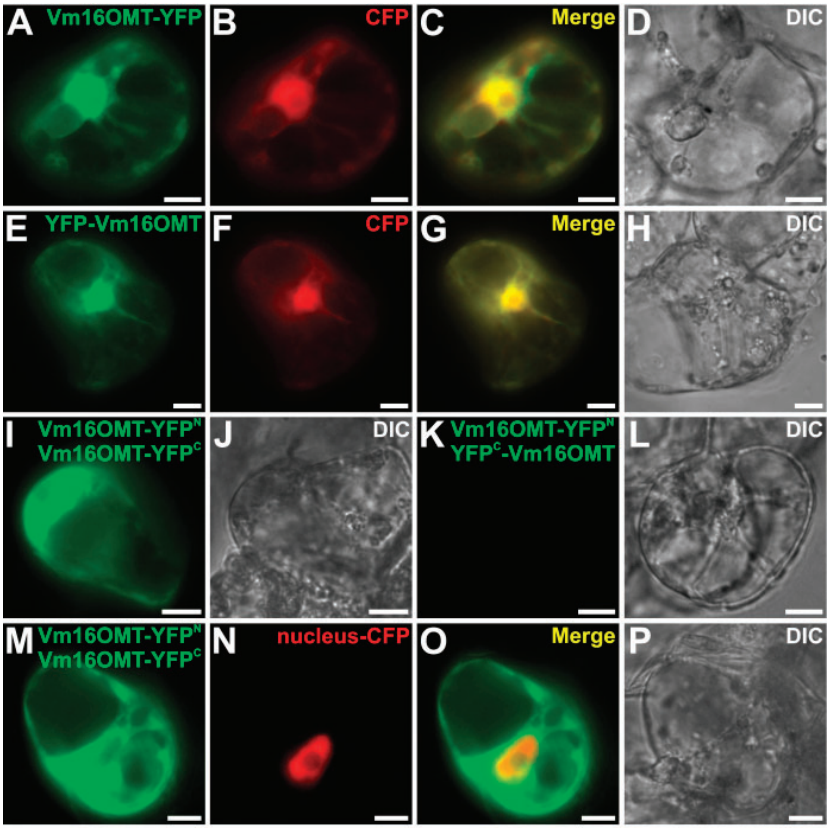


Figure 9. Subcellular localization of Vm16OMT in *C. roseus* cells. Vectors containing Vm16OMT-YFP (**A**–**D**) and YFP-Vm16OMT (**E**–**H**) were transiently co-expressed in *C. roseus* cells with a nucleocytoplasmic (CFP) marker. Bimolecular fluorescence complementation (BiFC) analyses (**I**–**P**) were performed by transiently co-expressing the corresponding constructs described in the upper right of each panel. A nuclear-CFP marker was included (**N**). Colocalization of the YFP and CFP fluorescence signals appear in the merged image. Cell morphology was observed using differential interference contrast (DIC). Bar 10 µm.

3.7. Gene Co-Expression Analysis

Finally, after validating our *V. minor* consensus transcriptome for homology-based identification of MIA genes, we took advantage of the multiple samples from the dataset to initiate the prediction of candidate genes catalyzing unknown MIA biosynthetic steps. Usually, genes sharing similar functions or involved in similar pathways display similar expression profiles [64]. Such similarities can be exploited in a "Guilt-By-Association" analysis to guide the identification of additional MIA pathway genes in the dataset [101]. For example, the downstream tabersonine pathway genes TEX and TAT were identified in *C. roseus* by searching candidates that were highly co-expressed with T16H [23,24]. Additionally, a co-expression analysis was used to short list candidates for subsequent *O*-methyltransferase activity assays towards identifying the 10-hydroxycamptothecin

O-methyltransferase (Ca10OMT) in *C. acuminata* [102]. Therefore, we conducted a co-expression analysis with the sequencing data generated from the four different tissue types (YL, YS, OL and AR). The predicted *V. minor* orthologs of MIA pathway genes from *C. roseus*, including the newly characterized Vm16OMT sequence, were used in the pathway level correlation (PLC) using pairwise Pearson's correlation coefficients (PCCs), ranked according to the highest reciprocal ranking (HRR) [87]. This analysis identified 39 co-expressed communities (Table S2). The four largest communities were found to be connected within a single large main network (Figure 10, communities 1, 2, 3 and 4).

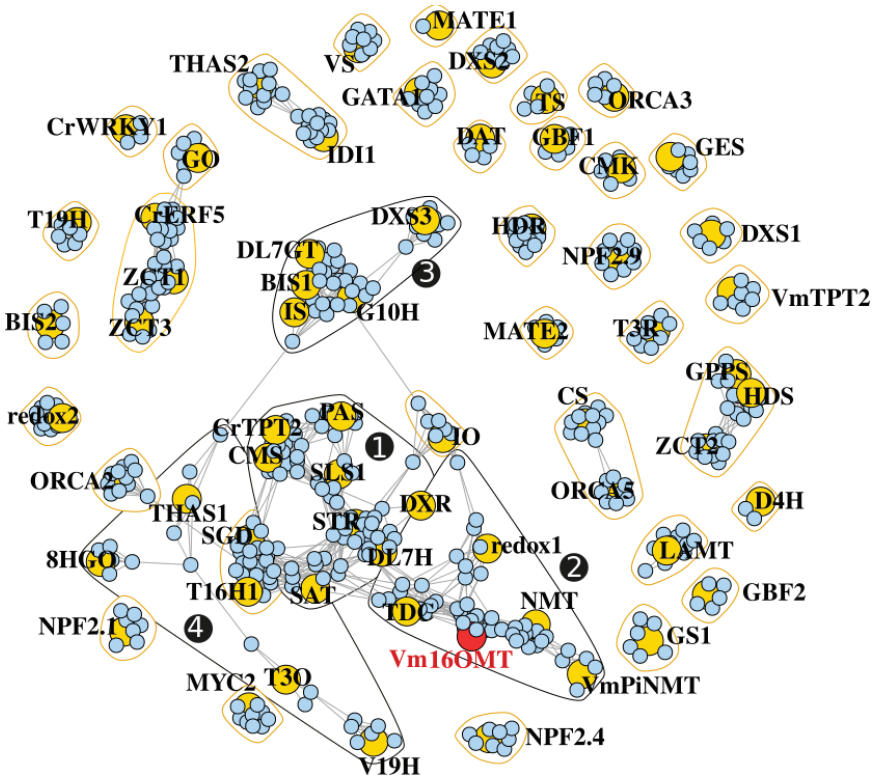


Figure 10. Co-expression network depicting the transcriptional relationship between annotated MIA genes (yellow) and *V. minor* transcripts (blue). Connected genes are indicated by grey edges for the retained HRR. Polygons correspond to communities detected using a fast greedy modularity optimization algorithm. The four main communities in the largest network are numbered with black circles and black polygons.

The largest group (community 1) revealed 63 transcripts upregulated with genes related to the seco-iridoid branch (CMS, DL7H, DXR and SLS) and STR, which catalyzes strictosidine biosynthesis in the first committed step towards MIA metabolism. This group also contained the downstream iboga- and aspidosperma-type MIA pathway genes responsible for the sequential conversion of stemmadenine into the central biosynthetic precursor, *O*-stemmadenine acetate (SAT and PAS). One of the co-expressed transcripts was predicted to encode a putative alcohol dehydrogenase and four were predicted to encode putative cytochrome P450s (Table S2).

The Vm16OMT was positioned within the second largest group (community 2) containing homologs to genes from the tabersonine/vincadifformine (redox1) and downstream vindoline pathways (NMT), co-expressed with a DXS homolog from the MEP pathway. Interestingly, an additional *O*-methyltransferase-like transcript was identified within this community, and two alcohol dehydrogenase-like transcripts (Table S2). Alcohol dehydrogenases and cytochrome P450's have long been implicated in MIA biosynthetic pathways where they catalyze numerous oxidation and reduction rearrangements of central precursor molecules, forming new types of scaffolds and decorations [46–48,103]. *V. minor* produces an assortment of MIA's containing such

decorations and rearrangements, potentially resulting from these identified candidates. For example, vincorine/norvincorine displays a methoxy on C15 [104], which probably involves a cytochrome P450 for hydroxylation, together with vincamine and ervinceine that both contain C17 and C16 hydroxyl groups, respectively [7,105]. Some of the associated P450 may thus constitute possible candidates for these unknown reactions.

The third group (community 3) contained genes related to the iridoid pathway (DL7GT, G10H and IS) co-expressed with DXR from the upstream MEP pathway, and the basic helix-loop-helix (bHLH) transcription factor, BIS2. BIS2 specifically regulates the iridoid branch of the pathway [106,107]. Capturing BIS2 tightly co-expressed with iridoid pathway genes thus added confidence to the network topology. However, no putative candidates can be identified from this third community and from the fourth one. This may illustrate that genes encoding enzymes from other branches of the pathway show a lower level of gene correlation. In agreement with this hypothesis, we identified orthologues of T16H and 2-hydroxyisoflavanone dehydratase in the fifth community that contains only nine correlated genes and no already known genes from the MIA pathway. In conclusion, our "Guilt-By-Association" analysis resulted in the identification of new candidate genes that can guide the future discovery of enzymes catalyzing hitherto unknown steps of MIA biosynthesis in *V. minor.*

4. Conclusions

In this study, we present a comprehensive and optimized *V. minor* consensus transcriptome compared to previous versions. This resource was used to predict MIA biosynthetic genes through a homology-based identification validated by the characterization of Vm16OMT. Gene correlation studies also resulted in the prediction of candidate genes putatively involved in unknown reactions of the pathway. As such, our work illustrates how gene discovery might be accelerated by using guilt-by-association studies together with the traditional homology-based approach. This also provides additional layers of evidence to aid in prioritizing MIA candidate genes before testing, and to guide detection of candidates that could have been missed due to low identity to known genes. In fact, it has already been shown that the closest gene homologs do not necessarily always correspond to true active candidates [93], highlighting the necessity for additional transcriptomic mining approaches for identification and prioritization of candidate genes. Based on this new resource, we can anticipate that the biosynthetic routes of vincamine and other less valuable MIAs will be elucidated in the near future. It will thus open new perspectives towards the development of microbial cell factories expressing the corresponding biosynthetic pathways for implementing an alternative supply of valuable MIAs.

Supplementary Materials: The following are available online at http://www.mdpi.com/2218-273X/10/12/1595/s1, Figure S1. Multidimensional scaling analysis of the RNA-seq experiments. Samples are projected on the two first dimensions according to their gene expression profiles. Closeness between samples on the grid indicates a strong similarity. Figure S2. Similar tissue types are strongly correlated. Spearman's rho correlation was measured to check biological replicate homogeneity. Higher rho values indicate stronger similarities between pairs of samples. Figure S3: DEG with logFC > 2 mapped to GO biological pathways that are significantly enriched ($p < 0.05$). Figure S4: DEG with logFC < 2 mapped GO biological pathways that are significantly enriched ($p < 0.05$). Figure S5: Significantly enriched GO biological pathways ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change < 2) YL vs. OL DEG comparison. Figure S6: Significantly enriched GO biological pathways ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change < 2) OL vs. AR DGE comparison. Figure S7: Significantly enriched GO biological pathways ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change < 2) YL vs. YS DEG comparison. Figure S8: Significantly enriched GO biological pathways ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change < 2) YL vs. AR DEG. Figure S9: Significantly enriched GO keywords ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change > 2) OL vs. AR DEG comparison. Figure S10: Significantly enriched GO keywords ($p < 0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change < 2) YL vs. OL DEG comparison. Figure S11. Comparison of expression levels (log2) obtained by qPCR and RNA-seq (transcripts per million) for candidate genes. A linear regression was used to model RNA-seq expression levels (expressed in transcripts per million) from qPCR values (normalized to actin raw expression). The blue line represents the regression line with a 95% confidence interval shown by the grey area. Figure S12: Alignment of the amino acid sequences of 16-*O*-tabersonine methyltransferase candidates. Sequence identity and similarity are highlighted by black and grey shading, respectively. The dimerization domain is highlighted in red and the *O*-methyltransferase domain is highlighted in blue. Figure S13: UPLC-DAD-MS analysis of 16-methoxytabersonine. (A) MS spectrum

and (B) UV spectrum. Table S1: Blastn analysis of *V. minor* consensus transcriptome against elucidated *C. roseus* MIA genes, MIA transporters and MIA transcription factors. GenBank accession numbers of *C. roseus* genes are displayed in brackets. Table S2: Transcripts that were co-expressed with MIA genes listed and arranged in communities that were determined by a fast greedy algorithm. Transcripts were annotated using blastx against the UniProt database. Table S3: Primers used in this study.

Author Contributions: Conceptualization, A.O., V.C.; methodology, E.A.S., L.J.S., T.D.d.B., I.C., K.K., P.L.C., S.B., A.L., N.P., N.G.-G., R.D., S.E.O., L.A., A.O., V.C.; software, E.A.S., L.J.S., T.D.d.B., R.D.; investigation, E.A.S., L.J.S., T.D.d.B., I.C., K.K., P.L.C., S.B., A.L., R.D, S.E.O.; resources, E.A.S., T.D.d.B., R.D.; data curation, E.A.S., T.D.d.B., R.D., S.B., A.L., V.C.; writing—original draft preparation, E.A.S., A.O., V.C.; writing—review and editing, E.A.S., A.O., V.C., N.P., N.G.-G., R.D., S.E.O., L.A., A.O., V.C.; supervision, V.C.; project administration, V.C.; funding acquisition, V.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the EU Horizon 2020 research and innovation program (MIAMi project-grant agreement N°814645), the ARD2020 Biopharmaceutical Program of the Région Centre Val de Loire (BioPROPHARM and CatharSIS projects), La Ligue contre le Cancer and Le Studium (Consortium fellowship).

Acknowledgments: We also thank Jean-Louis Rouet and Laurent Catherine for access and support to the CCSC computing resources (Cascimodot Federation, CNRS, Orléans).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Luijendijk, T.J.C.; van der Meijden, E.; Verpoorte, R. Involvement of Strictosidine as a Defensive Chemical in *Catharanthus roseus*. *J. Chem. Ecol.* **1996**, *22*, 1355. [CrossRef]
2. Roepke, J.; Salim, V.; Wu, M.; Thamm, A.M.K.; Murata, J.; Ploss, K.; Boland, W.; De Luca, V. Vinca Drug Components Accumulate Exclusively in Leaf Exudates of Madagascar Periwinkle. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15287–15292. [CrossRef]
3. Dugé De Bernonville, T.; Carqueijeiro, I.; Lanoue, A.; Lafontaine, F.; Sánchez Bel, P.; Liesecke, F.; Musset, K.; Oudin, A.; Glévarec, G.; Pichon, O.; et al. Folivory Elicits a Strong Defense Reaction in *Catharanthus roseus*: Metabolomic and Transcriptomic Analyses Reveal Distinct Local and Systemic Responses. *Sci. Rep.* **2017**, *7*, 1–14. [CrossRef] [PubMed]
4. Schläger, S.; Dräger, B. Exploiting Plant Alkaloids. *Curr. Opin. Biotechnol.* **2016**, *37*, 155–164. [CrossRef] [PubMed]
5. Carqueijeiro, I.; Langley, C.; Grzech, D.; Koudounas, K.; Papon, N.; O'Connor, S.E.; Courdavault, V. Beyond the Semi-Synthetic Artemisinin: Metabolic Engineering of Plant-Derived Anti-Cancer Drugs. *Curr. Opin. Biotechnol.* **2020**, *65*, 17–24. [CrossRef]
6. Courdavault, V.; O'Connor, S.E.; Oudin, A.; Besseau, S.; Papon, N. Towards the Microbial Production of Plant-Derived Anticancer Drugs. *Trends Cancer* **2020**, *6*, 444–448. [CrossRef] [PubMed]
7. Farahanikia, B.; Akbarzadeh, T.; Jahangirzadeh, A.; Yassa, N.; Shams Ardekani, M.R.; Mirnezami, T.; Hadjiakhoondi, A.; Khanavi, M. Phytochemical Investigation of *Vinca minor* Cultivated in Iran. *Iran. J. Pharm. Res. IJPR* **2011**, *10*, 777–785.
8. Demessie, Z.; Woolfson, K.N.; Yu, F.; Qu, Y.; De Luca, V. The ATP Binding Cassette Transporter, VmTPT2/VmABCG1, Is Involved in Export of the Monoterpenoid Indole Alkaloid, Vincamine in *Vinca minor* Leaves. *Phytochemistry* **2017**, *140*, 118–124. [CrossRef]
9. Grossmann, E.; Šefcővič, P.; Szász, K. Picrinine in *Vinca minor*. *Phytochemistry* **1973**, *12*, 2058. [CrossRef]
10. Verma, P.; Singh, N.; Khan, S.A.; Mathur, A.K.; Sharma, A.; Jamal, F. TIAs Pathway Genes and Associated MiRNA Identification in *Vinca minor*: Supporting Aspidosperma and Eburnamine Alkaloids Linkage via Transcriptomic Analysis. *Physiol. Mol. Biol. Plants* **2020**, *26*, 1695–1711. [CrossRef]
11. Levac, D.; Cázares, P.; Yu, F.; De Luca, V. A Picrinine *N*-Methyltransferase Belongs to a New Family of γ-Tocopherol-Like Methyltransferases Found in Medicinal Plants That Make Biologically Active Monoterpenoid Indole Alkaloids. *Plant Physiol.* **2016**, *170*, 1935. [CrossRef] [PubMed]
12. Abouzeid, S.; Beutling, U.; Surup, F.; Abdel Bar, F.M.; Amer, M.M.; Badria, F.A.; Yahyazadeh, M.; Brönstrup, M.; Selmar, D. Treatment of *Vinca minor* Leaves with Methyl Jasmonate Extensively Alters the Pattern and Composition of Indole Alkaloids. *J. Nat. Prod.* **2017**, *80*, 2905–2909. [CrossRef] [PubMed]
13. Murata, J.; Luca, V.D. Localization of Tabersonine 16-Hydroxylase and 16-OH Tabersonine-16-*O*-Methyltransferase to Leaf Epidermal Cells Defines Them as a Major Site of Precursor Biosynthesis in the Vindoline Pathway in *Catharanthus roseus*. *Plant J.* **2005**, *44*, 581–594. [CrossRef]

14. Murata, J.; Roepke, J.; Gordon, H.; De Luca, V. The Leaf Epidermome of *Catharanthus roseus* Reveals Its Biochemical Specialization. *Plant Cell* **2008**, *20*, 524. [CrossRef] [PubMed]
15. Levac, D.; Murata, J.; Kim, W.S.; De Luca, V. Application of Carborundum Abrasion for Investigating the Leaf Epidermis: Molecular Cloning of *Catharanthus roseus* 16-Hydroxytabersonine-16-*O*-Methyltransferase. *Plant J.* **2008**, *53*, 225–236. [CrossRef] [PubMed]
16. Proksa, B.; Grossmann, E. High Performance Liquid Chromatographic Determination of Alkaloids from *Vinca minor* L. *Phytochem. Anal.* **1991**, *2*, 74–76. [CrossRef]
17. O'Connor, S.E.; Maresh, J.J. Chemistry and Biology of Monoterpene Indole Alkaloid Biosynthesis. *Nat. Prod. Rep.* **2006**, *23*, 532–547. [CrossRef]
18. Salim, V.; De Luca, V. Chapter One-Towards Complete Elucidation of Monoterpene Indole Alkaloid Biosynthesis Pathway: *Catharanthus roseus* as a Pioneer System. In *Advances in Botanical Research*; Giglioli-Guivarc'h, N., Ed.; New Light on Alkaloid Biosynthesis and Future Prospects; Academic Press: Cambridge, MA, USA, 2013; Volume 68, pp. 1–37. [CrossRef]
19. Dugé de Bernonville, T.; Clastre, M.; Besseau, S.; Oudin, A.; Burlat, V.; Glévarec, G.; Lanoue, A.; Papon, N.; Giglioli-Guivarc'h, N.; St-Pierre, B.; et al. Phytochemical Genomics of the Madagascar Periwinkle: Unravelling the Last Twists of the Alkaloid Engine. *Phytochemistry* **2015**, *113*, 9–23. [CrossRef]
20. Thamm, A.M.K.; Qu, Y.; De Luca, V. Discovery and Metabolic Engineering of Iridoid/Secoiridoid and Monoterpenoid Indole Alkaloid Biosynthesis. *Phytochem. Rev.* **2016**, *15*, 339–361. [CrossRef]
21. Caputi, L.; Franke, J.; Farrow, S.C.; Chung, K.; Payne, R.M.E.; Nguyen, T.-D.; Dang, T.-T.T.; Soares Teto Carqueijeiro, I.; Koudounas, K.; Dugé de Bernonville, T.; et al. Missing Enzymes in the Biosynthesis of the Anticancer Drug Vinblastine in Madagascar Periwinkle. *Science* **2018**, *360*, 1235. [CrossRef] [PubMed]
22. Laflamme, P.; St-Pierre, B.; De Luca, V. Molecular and Biochemical Analysis of a Madagascar Periwinkle Root-Specific Minovincinine-19-Hydroxy-*O*-Acetyltransferase. *Plant Physiol.* **2001**, *125*, 189. [CrossRef] [PubMed]
23. Carqueijeiro, I.; Brown, S.; Chung, K.; Dang, T.-T.; Walia, M.; Besseau, S.; Dugé de Bernonville, T.; Oudin, A.; Lanoue, A.; Billet, K.; et al. Two Tabersonine 6,7-Epoxidases Initiate Lochnericine-Derived Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant Physiol.* **2018**, *177*, 1473. [CrossRef] [PubMed]
24. Carqueijeiro, I.; Dugé de Bernonville, T.; Lanoue, A.; Dang, T.-T.; Teijaro, C.N.; Paetz, C.; Billet, K.; Mosquera, A.; Oudin, A.; Besseau, S.; et al. A BAHD Acyltransferase Catalyzing 19-*O*-Acetylation of Tabersonine Derivatives in Roots of *Catharanthus roseus* Enables Combinatorial Synthesis of Monoterpene Indole Alkaloids. *Plant J.* **2018**, *94*, 469–484. [CrossRef]
25. Williams, D.; Qu, Y.; Simionescu, R.; De Luca, V. The Assembly of (+)-Vincadifformine- and (−)-Tabersonine-Derived Monoterpenoid Indole Alkaloids in *Catharanthus roseus* Involves Separate Branch Pathways. *Plant J.* **2019**, *99*, 626–636. [CrossRef]
26. Kutchan, T.M.; Hampp, N.; Lottspeich, F.; Beyreuther, K.; Zenk, M.H. The CDNA Clone for Strictosidine Synthase from *Rauvolfia serpentina* DNA Sequence Determination and Expression in *Escherichia coli*. *FEBS Lett.* **1988**, *237*, 40–44. [CrossRef]
27. McKnight, T.D.; Roessner, C.A.; Devagupta, R.; Scott, A.I.; Nessler, C.L. Nucleotide Sequence of a CDNA Encoding the Vacuolar Protein Strictosidine Synthase from *Catharanthus roseus*. *Nucleic Acids Res.* **1990**, *18*, 4939. [CrossRef]
28. Kellner, F.; Geu-Flores, F.; Sherden, N.H.; Brown, S.; Foureau, E.; Courdavault, V.; O'Connor, S.E. Discovery of a P450-Catalyzed Step in Vindoline Biosynthesis: A Link between the Aspidosperma and Eburnamine Alkaloids. *Chem. Commun.* **2015**, *51*, 7626–7628. [CrossRef]
29. De Luca, V.; Marineau, C.; Brisson, N. Molecular Cloning and Analysis of CDNA Encoding a Plant Tryptophan Decarboxylase: Comparison with Animal Dopa Decarboxylases. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2582. [CrossRef]
30. Goddijn, O.J.M.; Lohman, F.P.; de Kam, R.J.; hilperoort, R.A.; Hoge, J.H.C. Nucleotide Sequence of the Tryptophan Decarboxylase Gene of *Catharanthus roseus* and Expression of Tdc-GusA Gene Fusions in *Nicotiana tabacum*. *Mol. Gen. Genet. MGG* **1994**, *242*, 217–225. [CrossRef]
31. Veau, B.; Courtois, M.; Oudin, A.; Chénieux, J.-C.; Rideau, M.; Clastre, M. Cloning and Expression of CDNAs Encoding Two Enzymes of the MEP Pathway in *Catharanthus roseus*. *Biochim. Biophys. Acta* **2000**, *1517*, 159–163. [CrossRef]

32. Oudin, A.; Mahroug, S.; Courdavault, V.; Hervouet, N.; Zelwer, C.; Rodríguez-Concepción, M.; St-Pierre, B.; Burlat, V. Spatial Distribution and Hormonal Regulation of Gene Products from Methyl Erythritol Phosphate and Monoterpene-Secoiridoid Pathways in *Catharanthus roseus*. *Plant Mol. Biol.* **2007**, *65*, 13–30. [CrossRef] [PubMed]
33. Guirimand, G.; Guihur, A.; Phillips, M.A.; Oudin, A.; Glévarec, G.; Melin, C.; Papon, N.; Clastre, M.; St-Pierre, B.; Rodríguez-Concepción, M.; et al. A Single Gene Encodes Isopentenyl Diphosphate Isomerase Isoforms Targeted to Plastids, Mitochondria and Peroxisomes in *Catharanthus roseus*. *Plant Mol. Biol.* **2012**, *79*, 443–459. [CrossRef] [PubMed]
34. Rai, A.; Smita, S.S.; Singh, A.K.; Shanker, K.; Nagegowda, D.A. Heteromeric and Homomeric Geranyl Diphosphate Synthases from *Catharanthus roseus* and Their Role in Monoterpene Indole Alkaloid Biosynthesis. *Mol. Plant* **2013**, *6*, 1531–1549. [CrossRef] [PubMed]
35. Guirimand, G.; Guihur, A.; Perello, C.; Phillips, M.; Mahroug, S.; Oudin, A.; Dugé de Bernonville, T.; Besseau, S.; Lanoue, A.; Giglioli-Guivarc'h, N.; et al. Cellular and Subcellular Compartmentation of the 2C-Methyl-D-Erythritol 4-Phosphate Pathway in the Madagascar Periwinkle. *Plants* **2020**, *9*, 462. [CrossRef] [PubMed]
36. Simkin, A.J.; Miettinen, K.; Claudel, P.; Burlat, V.; Guirimand, G.; Courdavault, V.; Papon, N.; Meyer, S.; Godet, S.; St-Pierre, B.; et al. Characterization of the Plastidial Geraniol Synthase from Madagascar Periwinkle Which Initiates the Monoterpenoid Branch of the Alkaloid Pathway in Internal Phloem Associated Parenchyma. *Phytochemistry* **2013**, *85*, 36–43. [CrossRef]
37. Collu, G.; Unver, N.; Peltenburg-Looman, A.M.G.; van der Heijden, R.; Verpoorte, R.; Memelink, J. Geraniol 10-Hydroxylase1, a Cytochrome P450 Enzyme Involved in Terpenoid Indole Alkaloid Biosynthesis. *FEBS Lett.* **2001**, *508*, 215–220. [CrossRef]
38. Miettinen, K.; Dong, L.; Navrot, N.; Schneider, T.; Burlat, V.; Pollier, J.; Woittiez, L.; van der Krol, S.; Lugan, R.; Ilc, T.; et al. The Seco-Iridoid Pathway from *Catharanthus roseus*. *Nat. Commun.* **2014**, *5*, 3606. [CrossRef]
39. Geu-Flores, F.; Sherden, N.H.; Courdavault, V.; Burlat, V.; Glenn, W.S.; Wu, C.; Nims, E.; Cui, Y.; O'Connor, S.E. An Alternative Route to Cyclic Terpenes by Reductive Cyclization in Iridoid Biosynthesis. *Nature* **2012**, *492*, 138–142. [CrossRef]
40. Salim, V.; Wiens, B.; Masada-Atsumi, S.; Yu, F.; De Luca, V. 7-Deoxyloganetic Acid Synthase Catalyzes a Key 3 Step Oxidation to Form 7-Deoxyloganetic Acid in *Catharanthus roseus* Iridoid Biosynthesis. *Phytochemistry* **2014**, *101*, 23–31. [CrossRef]
41. Asada, K.; Salim, V.; Masada-Atsumi, S.; Edmunds, E.; Nagatoshi, M.; Terasaka, K.; Mizukami, H.; De Luca, V. A 7-Deoxyloganetic Acid Glucosyltransferase Contributes a Key Step in Secologanin Biosynthesis in Madagascar Periwinkle. *Plant Cell* **2013**, *25*, 4123. [CrossRef] [PubMed]
42. Salim, V.; Yu, F.; Altarejos, J.; De Luca, V. Virus-Induced Gene Silencing Identifies *Catharanthus roseus* 7-Deoxyloganic Acid-7-Hydroxylase, a Step in Iridoid and Monoterpene Indole Alkaloid Biosynthesis. *Plant J.* **2013**, *76*, 754–765. [CrossRef] [PubMed]
43. Irmler, S.; Schröder, G.; St-Pierre, B.; Crouch, N.P.; Hotze, M.; Schmidt, J.; Strack, D.; Matern, U.; Schröder, J. Indole Alkaloid Biosynthesis in *Catharanthus roseus*: New Enzyme Activities and Identification of Cytochrome P450 CYP72A1 as Secologanin Synthase. *Plant J.* **2000**, *24*, 797–804. [CrossRef] [PubMed]
44. Luijendijk, T.J.C.; Stevens, L.H.; Verpoorte, R. Purification and Characterisation of Strictosidine β-d-Glucosidase from *Catharanthus roseus* Cell Suspension Cultures. *Plant Physiol. Biochem.* **1998**, *36*, 419–425. [CrossRef]
45. Geerlings, A.; Ibañez, M.M.-L.; Memelink, J.; van der Heijden, R.; Verpoorte, R. Molecular Cloning and Analysis of Strictosidine β-d-Glucosidase, an Enzyme in Terpenoid Indole Alkaloid Biosynthesis in *Catharanthus roseus*. *J. Biol. Chem.* **2000**, *275*, 3051–3056. [CrossRef] [PubMed]
46. Stavrinides, A.; Tatsis, E.C.; Foureau, E.; Caputi, L.; Kellner, F.; Courdavault, V.; O'Connor, S.E. Unlocking the Diversity of Alkaloids in *Catharanthus roseus*: Nuclear Localization Suggests Metabolic Channeling in Secondary Metabolism. *Chem. Biol.* **2015**, *22*, 336–341. [CrossRef]
47. Stavrinides, A.; Tatsis, E.C.; Caputi, L.; Foureau, E.; Stevenson, C.E.M.; Lawson, D.M.; Courdavault, V.; O'Connor, S.E. Structural Investigation of Heteroyohimbine Alkaloid Synthesis Reveals Active Site Elements That Control Stereoselectivity. *Nat. Commun.* **2016**, *7*, 12116. [CrossRef]

48. Tatsis, E.C.; Carqueijeiro, I.; Dugé de Bernonville, T.; Franke, J.; Dang, T.-T.T.; Oudin, A.; Lanoue, A.; Lafontaine, F.; Stavrinides, A.K.; Clastre, M.; et al. A Three Enzyme System to Generate the Strychnos Alkaloid Scaffold from a Central Biosynthetic Intermediate. *Nat. Commun.* **2017**, *8*, 316. [CrossRef]
49. Qu, Y.; Thamm, A.M.K.; Czerwinski, M.; Masada, S.; Kim, K.H.; Jones, G.; Liang, P.; De Luca, V. Geissoschizine Synthase Controls Flux in the Formation of Monoterpenoid Indole Alkaloids in a *Catharanthus roseus* Mutant. *Planta* **2018**, *247*, 625–634. [CrossRef]
50. Qu, Y.; Easson, M.L.A.E.; Froese, J.; Simionescu, R.; Hudlicky, T.; De Luca, V. Completion of the Seven-Step Pathway from Tabersonine to the Anticancer Drug Precursor Vindoline and Its Assembly in Yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6224. [CrossRef]
51. St-Pierre, B.; De Luca, V. A Cytochrome P-450 Monooxygenase Catalyzes the First Step in the Conversion of Tabersonine to Vindoline in *Catharanthus roseus*. *Plant Physiol.* **1995**, *109*, 131. [CrossRef] [PubMed]
52. Schröder, G.; Unterbusch, E.; Kaltenbach, M.; Schmidt, J.; Strack, D.; De Luca, V.; Schröder, J. Light-Induced Cytochrome P450-Dependent Enzyme in Indole Alkaloid Biosynthesis: Tabersonine 16-Hydroxylase. *FEBS Lett.* **1999**, *458*, 97–102. [CrossRef]
53. Besseau, S.; Kellner, F.; Lanoue, A.; Thamm, A.M.K.; Salim, V.; Schneider, B.; Geu-Flores, F.; Höfer, R.; Guirimand, G.; Guihur, A.; et al. A Pair of Tabersonine 16-Hydroxylases Initiates the Synthesis of Vindoline in an Organ-Dependent Manner in *Catharanthus roseus*. *Plant Physiol.* **2013**, *163*, 1792. [CrossRef]
54. De Luca, V.; Fernandez, J.A.; Campbell, D.; Kurz, W.G.W. Developmental Regulation of Enzymes of Indole Alkaloid Biosynthesis in *Catharanthus roseus*. *Plant Physiol.* **1988**, *86*, 447–450. [CrossRef] [PubMed]
55. Liscombe, D.K.; Usera, A.R.; O'Connor, S.E. Homolog of Tocopherol *C* Methyltransferases Catalyzes *N* Methylation in Anticancer Alkaloid Biosynthesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18793. [CrossRef]
56. Vazquez-Flota, F.; De Carolis, E.; Alarco, A.M.; De Luca, V. Molecular Cloning and Characterization of Desacetoxyvindoline-4-Hydroxylase, a 2-Oxoglutarate Dependent-Dioxygenase Involved in the Biosynthesis of Vindoline in *Catharanthus roseus* (L.) G. Don. *Plant Mol. Biol.* **1997**, *34*, 935–948. [CrossRef] [PubMed]
57. St-Pierre, B.; Laflamme, P.; Alarco, A.-M.; De Luca, V. The Terminal *O*-Acetyltransferase Involved in Vindoline Biosynthesis Defines a New Class of Proteins Responsible for Coenzyme A-Dependent Acyl Transfer. *Plant J.* **1998**, *14*, 703–713. [CrossRef]
58. De Luca, V.; Balsevich, J.; Tyler, R.T.; Eilert, U.; Panchuk, B.D.; Kurz, W.G.W. Biosynthesis of Indole Alkaloids: Developmental Regulation of the Biosynthetic Pathway from Tabersonine to Vindoline in *Catharanthus roseus*. *J. Plant Physiol.* **1986**, *125*, 147–156. [CrossRef]
59. De Carolis, E.; Chan, F.; Balsevich, J.; De Luca, V. Isolation and Characterization of a 2-Oxoglutarate Dependent Dioxygenase Involved in the Second-to-Last Step in Vindoline Biosynthesis. *Plant Physiol.* **1990**, *94*, 1323–1329. [CrossRef]
60. Giddings, L.-A.; Liscombe, D.K.; Hamilton, J.P.; Childs, K.L.; DellaPenna, D.; Buell, C.R.; O'Connor, S.E. A Stereoselective Hydroxylation Step of Alkaloid Biosynthesis by a Unique Cytochrome P450 in *Catharanthus roseus*. *J. Biol. Chem.* **2011**, *286*, 16751–16757. [CrossRef]
61. Chen, Q.; Lu, X.; Guo, X.; Guo, Q.; Li, D. Metabolomics Characterization of Two Apocynaceae Plants, *Catharanthus roseus* and *Vinca minor*, Using GC-MS and LC-MS Methods in Combination. *Molecules* **2017**, *22*, 997. [CrossRef] [PubMed]
62. Wenkert, E.; Wickberg, B. General Methods of Synthesis of Indole Alkaloids. IV. A Synthesis of Dl-Eburnamonine1, 2. *J. Am. Chem. Soc.* **1965**, *87*, 1580–1589. [CrossRef] [PubMed]
63. Kries, H.; O'Connor, S.E. Biocatalysts from Alkaloid Producing Plants. *Curr. Opin. Chem. Biol.* **2016**, *31*, 22–30. [CrossRef] [PubMed]
64. Dugé de Bernonville, T.; Papon, N.; Clastre, M.; O'Connor, S.E.; Courdavault, V. Identifying Missing Biosynthesis Enzymes of Plant Natural Products. *Trends Pharmacol. Sci.* **2020**, *41*, 142–146. [CrossRef] [PubMed]
65. Parage, C.; Foureau, E.; Kellner, F.; Burlat, V.; Mahroug, S.; Lanoue, A.; de Bernonville, T.D.; Londono, M.A.; Carqueijeiro, I.; Oudin, A.; et al. Class II Cytochrome P450 Reductase Governs the Biosynthesis of Alkaloids. *Plant Physiol.* **2016**, *172*, 1563–1577. [CrossRef] [PubMed]
66. Brown, S.; Clastre, M.; Courdavault, V.; O'Connor, S.E. De Novo Production of the Plant-Derived Alkaloid Strictosidine in Yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3205–3210. [CrossRef]

67. Pyne, M.E.; Kevvai, K.; Grewal, P.S.; Narcross, L.; Choi, B.; Bourgeois, L.; Dueber, J.E.; Martin, V.J.J. A Yeast Platform for High-Level Synthesis of Tetrahydroisoquinoline Alkaloids. *Nat. Commun.* **2020**, *11*, 3337. [CrossRef]
68. Zhao, S.; Andrade, R.B. Domino Michael/Mannich/*N*-Alkylation Route to the Tetrahydrocarbazole Framework of Aspidosperma Alkaloids: Concise Total Syntheses of (−)-Aspidospermidine, (−)-Tabersonine, and (−)-Vincadifformine. *J. Am. Chem. Soc.* **2013**, *135*, 13334–13337. [CrossRef]
69. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
70. Chang, Z.; Li, G.; Liu, J.; Zhang, Y.; Ashby, C.; Liu, D.; Cramer, C.L.; Huang, X. Bridger: A New Framework for de Novo Transcriptome Assembly Using RNA-Seq Data. *Genome Biol.* **2015**, *16*, 30. [CrossRef]
71. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef]
72. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef]
73. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Methods in Molecular Biology*; Humana Press Inc.: Totowa, NJ, USA, 2019; Volume 1962, pp. 227–245. [CrossRef]
74. Bryant, D.M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M.B.; Payzin-Dogru, D.; Lee, T.J.; Leigh, N.D.; Kuo, T.-H.; Davis, F.G.; et al. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep.* **2017**, *18*, 762–776. [CrossRef]
75. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. *De Novo* Transcript Sequence Reconstruction from RNA-Seq: Reference Generation and Analysis with Trinity. *Nat. Protoc.* **2013**, *8*. [CrossRef]
76. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]
77. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2009**, *26*, 139–140. [CrossRef]
78. Conway, J.R.; Lex, A.; Gehlenborg, N. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* **2017**, *33*, 2938–2940. [CrossRef]
79. R Core Team. *R: A Language and Environment for Statistical Computing*; Vienna, Austria, 2013. Available online: https://www.r-project.org/ (accessed on 24 November 2020).
80. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef]
81. Wickham, H.G. *gplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2009.
82. Guirimand, G.; Burlat, V.; Oudin, A.; Lanoue, A.; St-Pierre, B.; Courdavault, V. Optimization of the Transient Transformation of Catharanthus roseus Cells by Particle Bombardment and Its Application to the Subcellular Localization of Hydroxymethylbutenyl 4-Diphosphate Synthase and Geraniol 10-Hydroxylase. *Plant Cell Rep.* **2009**, *28*, 1215–1234. [CrossRef] [PubMed]
83. Waadt, R.; Kudla, J. In Planta Visualization of Protein Interactions Using Bimolecular Fluorescence Complementation (BiFC). *Cold Spring Harb. Protoc.* **2008**, *2008*, pdb.prot4995. [CrossRef] [PubMed]
84. Guirimand, G.; Guihur, A.; Ginis, O.; Poutrain, P.; Héricourt, F.; Oudin, A.; Lanoue, A.; St-Pierre, B.; Burlat, V.; Courdavault, V. The Subcellular Organization of Strictosidine Biosynthesis in *Catharanthus roseus* Epidermis Highlights Several Trans-Tonoplast Translocations of Intermediate Metabolites. *FEBS J.* **2011**, *278*, 749–763. [CrossRef] [PubMed]
85. Foureau, E.; Carqueijeiro, I.; Dugé de Bernonville, T.; Melin, C.; Lafontaine, F.; Besseau, S.; Lanoue, A.; Papon, N.; Oudin, A.; Glévarec, G.; et al. Chapter Eight-Prequels to Synthetic Biology: From Candidate Gene Identification and Validation to Enzyme Subcellular Localization in Plant and Yeast Cells. In *Methods in Enzymology*; O'Connor, S.E., Ed.; Synthetic Biology and Metabolic Engineering in Plants and Microbes Part B: Metabolism in Plants; Academic Press: Cambridge, MA, USA, 2016; Volume 576, pp. 167–206. [CrossRef]
86. Mutwil, M.; Klie, S.; Tohge, T.; Giorgi, F.M.; Wilkins, O.; Campbell, M.M.; Fernie, A.R.; Usadel, B.; Nikoloski, Z.; Persson, S. PlaNet: Combined Sequence and Expression Comparisons across Plant Networks Derived from Seven Species. *Plant Cell* **2011**, *23*, 895–910. [CrossRef] [PubMed]

87. Liesecke, F.; Daudu, D.; Dugé de Bernonville, R.; Besseau, S.; Clastre, M.; Courdavault, V.; de Craene, J.-O.; Crèche, J.; Giglioli-Guivarc'h, N.; Glévarec, G.; et al. Ranking Genome-Wide Correlation Measurements Improves Microarray and RNA-Seq Based Global and Targeted Co-Expression Networks. *Sci. Rep.* **2018**, *8*, 10885. [CrossRef]
88. Csárdi, G.; Nepusz, T. The Igraph Software Package for Complex. Network Research. Available online: researchgate.net (accessed on 18 November 2020).
89. Clauset, A.; Newman, M.E.J.; Moore, C. Finding Community Structure in Very Large Networks. *Phys. Rev. E* **2004**, *70*, 66111. [CrossRef] [PubMed]
90. Matsuura, H.N.; Fragoso, V.; Paranhos, J.T.; Rau, M.R.; Fett-Neto, A.G. The Bioactive Monoterpene Indole Alkaloid N,β-d-Glucopyranosyl Vincosamide Is Regulated by Irradiance Quality and Development in *Psychotria leiocarpa*. *Ind. Crops Prod.* **2016**, *86*, 210–218. [CrossRef]
91. Liu, Y.; Song, L.; Yu, W.; Hu, Y.; Ma, X.; Wu, J.; Ying, Y. Light Quality Modifies Camptothecin Production and Gene Expression of Biosynthesis in *Camptotheca acuminata* Decne Seedlings. *Ind. Crops Prod.* **2015**, *66*, 137–143. [CrossRef]
92. Xiao, M.; Zhang, Y.; Chen, X.; Lee, E.J.; Barber, C.J.S.; Chakrabarty, R.; Desgagné-Penix, I.; Haslam, T.M.; Kim, Y.B.; Liu, E.; et al. Transcriptome Analysis Based on Next-Generation Sequencing of Non-Model Plants Producing Specialized Metabolites of Biotechnological Interest. *J. Biotechnol.* **2013**, *166*, 122–134. [CrossRef]
93. Franke, J.; Kim, J.; Hamilton, J.P.; Zhao, D.; Pham, G.M.; Wiegert-Rininger, K.; Crisovan, E.; Newton, L.; Vaillancourt, B.; Tatsis, E.; et al. Gene Discovery in Gelsemium Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis. *ChemBioChem* **2019**, *20*, 83–87. [CrossRef]
94. Zhao, D.; Hamilton, J.P.; Pham, G.M.; Crisovan, E.; Wiegert-Rininger, K.; Vaillancourt, B.; DellaPenna, D.; Robin Buell, C. *De Novo* Genome Assembly of *Camptotheca acuminata*, a Natural Source of the Anti-Cancer Compound Camptothecin. *GigaScience* **2017**, *6*, 1–7. [CrossRef] [PubMed]
95. Courdavault, V.; Papon, N.; Clastre, M.; Giglioli-Guivarc'h, N.; St-Pierre, B.; Burlat, V. A Look inside an Alkaloid Multisite Plant: The Catharanthus Logistics. *Curr. Opin. Plant Biol.* **2014**, *19*, 43–50. [CrossRef] [PubMed]
96. Proksa, B.; Uhrín, D.; Grossmann, E.; Votický, Z.; Fuska, J. Relative Configuration and Cytotoxic Activity of Vincarubine: A Novel Bisindole Alkaloid from *Vinca minor*. *Planta Med.* **1988**, *54*, 214–218. [CrossRef]
97. Lichman, B.R.; Godden, G.T.; Buell, C.R. Gene and Genome Duplications in the Evolution of Chemodiversity: Perspectives from Studies of Lamiaceae. *Curr. Opin. Plant. Biol.* **2020**, *55*, 74–83. [CrossRef] [PubMed]
98. Duplais, C.; Papon, N.; Courdavault, V. Tracking the Origin and Evolution of Plant Metabolites. *Trends Plant Sci.* **2020**. [CrossRef] [PubMed]
99. Zubieta, C.; He, X.-Z.; Dixon, R.A.; Noel, J.P. Structures of Two Natural Product Methyltransferases Reveal the Basis for Substrate Specificity in Plant *O*-Methyltransferases. *Nat. Struct. Biol.* **2001**, *8*, 271–279. [CrossRef] [PubMed]
100. Weis, K. Regulating Access to the Genome: Nucleocytoplasmic Transport throughout the Cell Cycle. *Cell* **2003**, *112*, 441–451. [CrossRef]
101. Oliver, S. Guilt-by-association goes global. *Nature* **2000**, *403*, 601–603. [CrossRef]
102. Salim, V.; Jones, A.D.; DellaPenna, D. *Camptotheca acuminata* 10-Hydroxycamptothecin *O*-Methyltransferase: An Alkaloid Biosynthetic Enzyme Co-Opted from Flavonoid Metabolism. *Plant J.* **2018**, *95*, 112–125. [CrossRef]
103. Werck-Reichhart, D.; Feyereisen, R. Cytochromes P450: A Success Story. *Genome Biol.* **2000**, *1*, 1–9. [CrossRef]
104. Eckermann, R.; Gaich, T. The Akuammiline Alkaloids; Origin and Synthesis. *Synthesis* **2013**, *45*, 2813–2823. [CrossRef]
105. Rakhimov, D.A.; Malikov, V.M.; Yagudaev, M.R.; Yunusov, S.Y. Structures of Ervinceine, Ervamicine, and Ervincinine. *Chem. Nat. Compd.* **1970**, *6*, 221–224. [CrossRef]
106. Van Moerkercke, A.; Steensma, P.; Schweizer, F.; Pollier, J.; Gariboldi, I.; Payne, R.; Vanden Bossche, R.; Miettinen, K.; Espoz, J.; Purnama, P.C.; et al. The BHLH Transcription Factor BIS1 Controls the Iridoid Branch of the Monoterpenoid Indole Alkaloid Pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8130–8135. [CrossRef] [PubMed]

107. Van Moerkercke, A.; Steensma, P.; Gariboldi, I.; Espoz, J.; Purnama, P.C.; Schweizer, F.; Miettinen, K.; Vanden Bossche, R.; De Clercq, R.; Memelink, J.; et al. The Basic Helix-Loop-Helix Transcription Factor BIS2 Is Essential for Monoterpenoid Indole Alkaloid Production in the Medicinal Plant *Catharanthus roseus*. *Plant J.* **2016**, *88*, 3–12. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

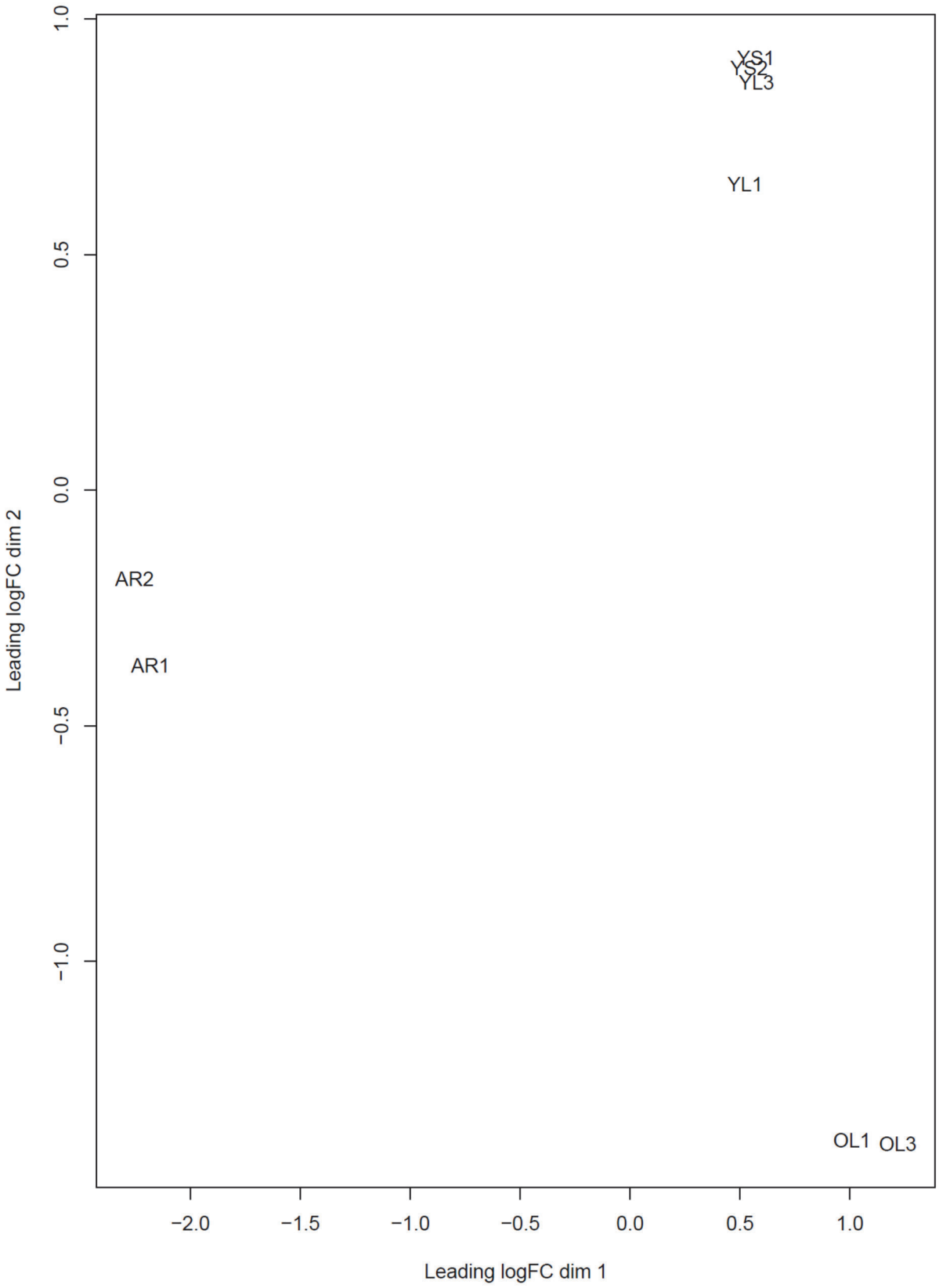


Figure S1. Multidimensional scaling analysis of the RNA-seq experiments. Samples are projected on the two first dimensions according to their gene expression profiles. Closeness between samples on the grid indicates a strong similarity.

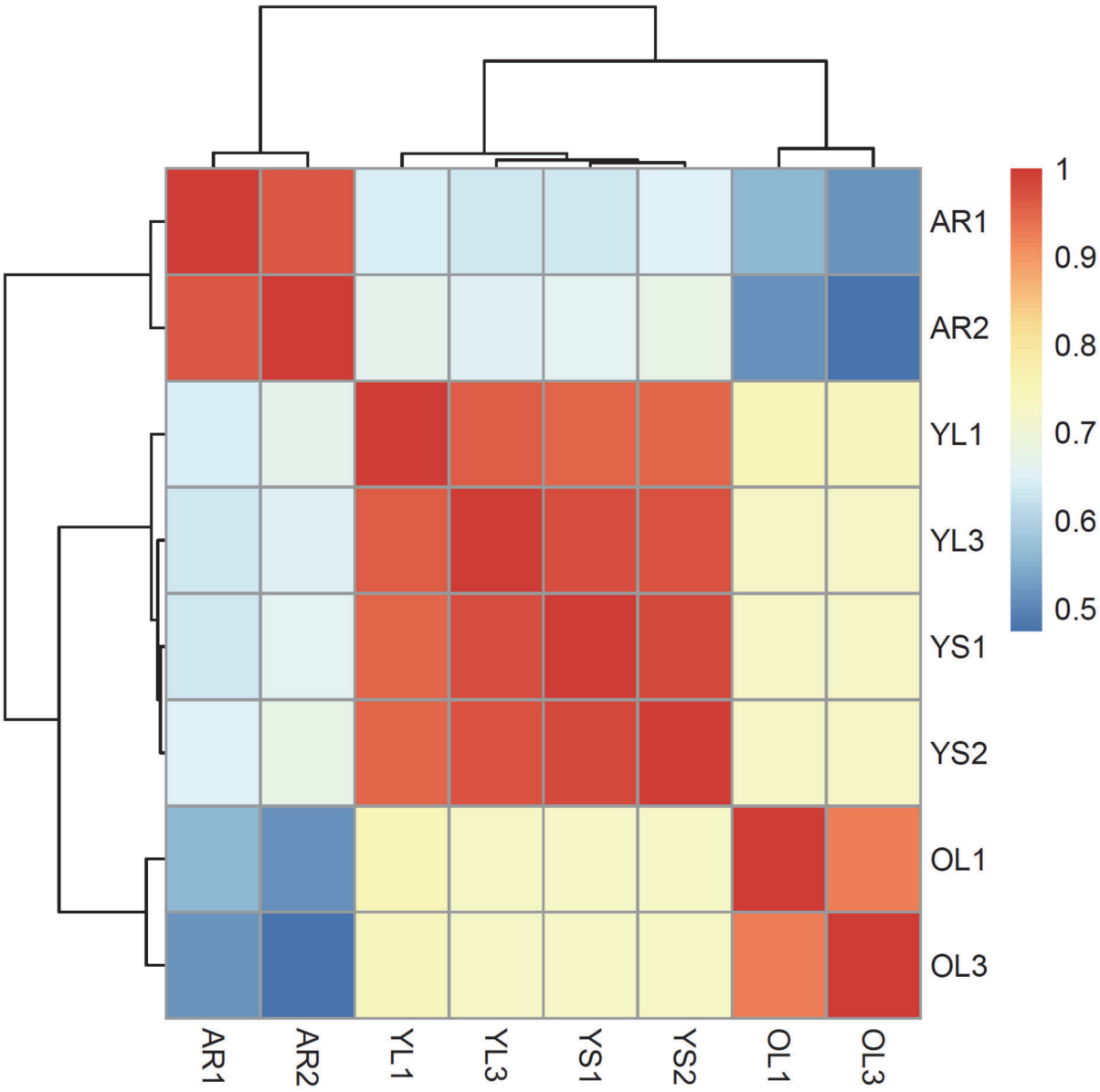


Figure S2. Similar tissue type are strongly correlated. Spearman's rho correlation was measured to check biological replicate homogeneity. Higher rho values indicate stronger similarities between pairs of samples.

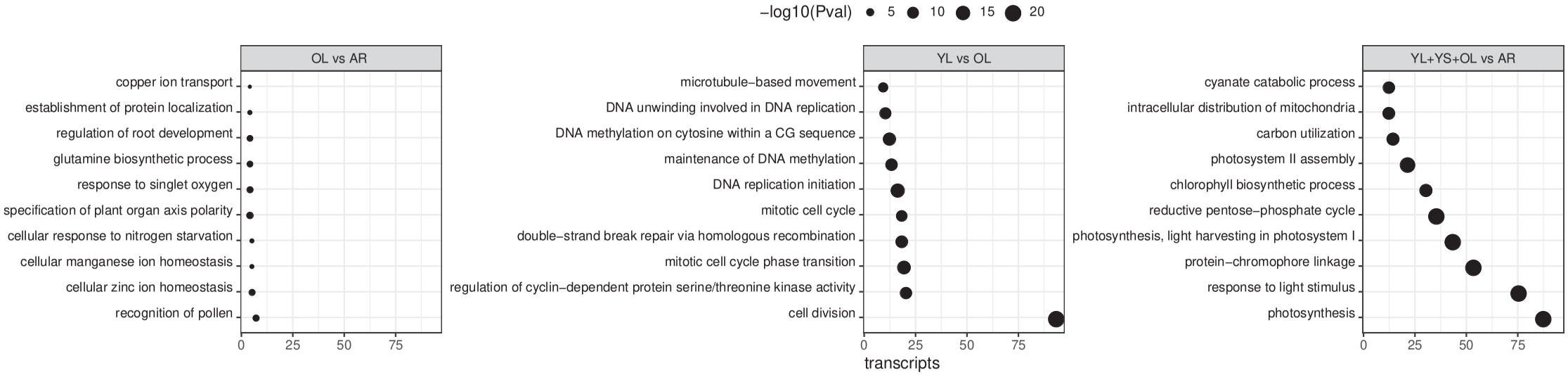


Figure S3. DEG with logFC > 2 mapped to GO biological pathways that are significantly enriched (p<0.05)

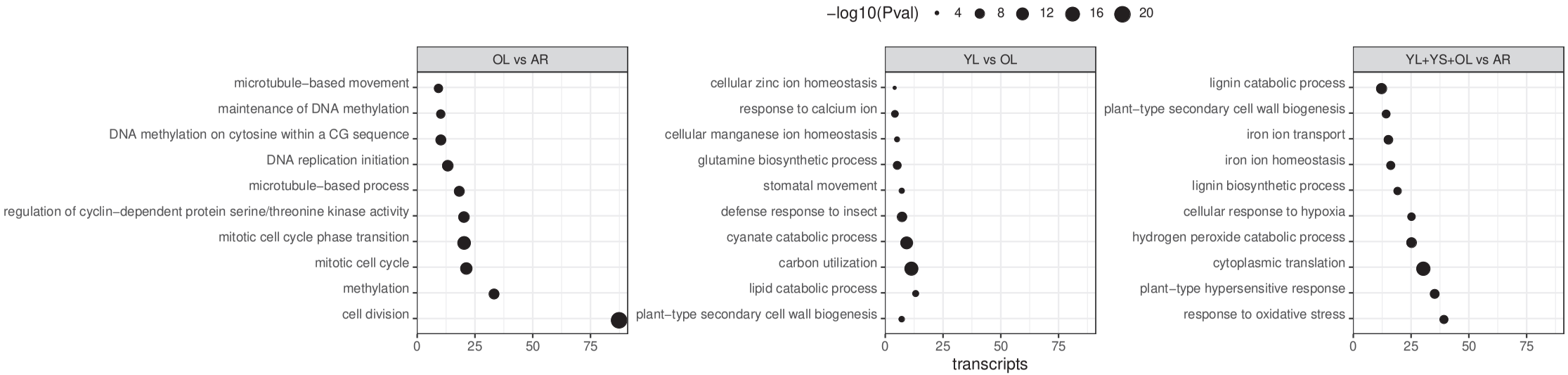


Figure S4. DEG with logFC < 2 mapped GO biological pathways that are significantly enriched (p<0.05).

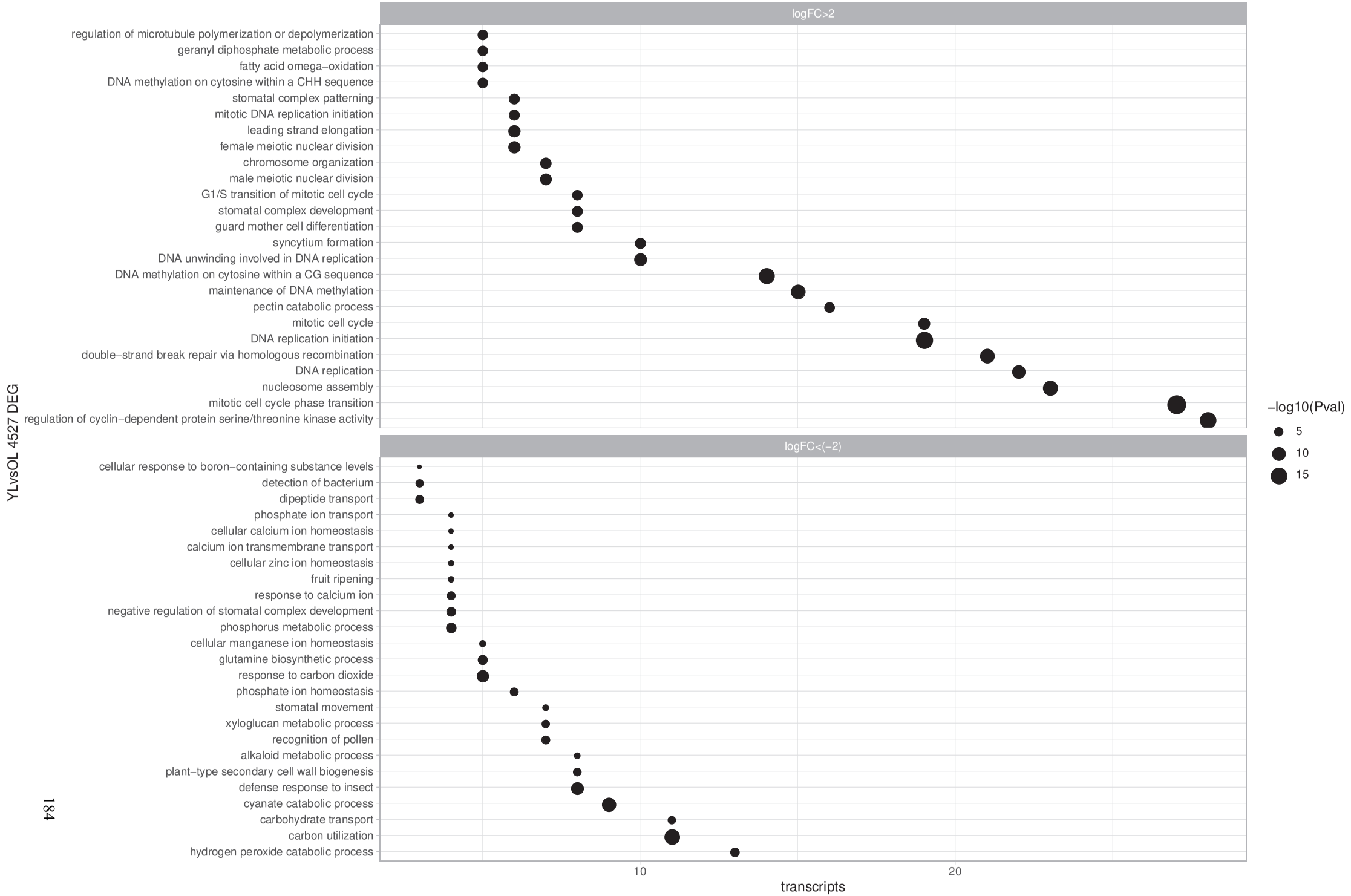


Figure S5. Significantly enriched GO biological pathways (p<0.05) in the upregulated (log fold change > 2) and downregulated (log fold change <2) YLvsOL DEG comparison.

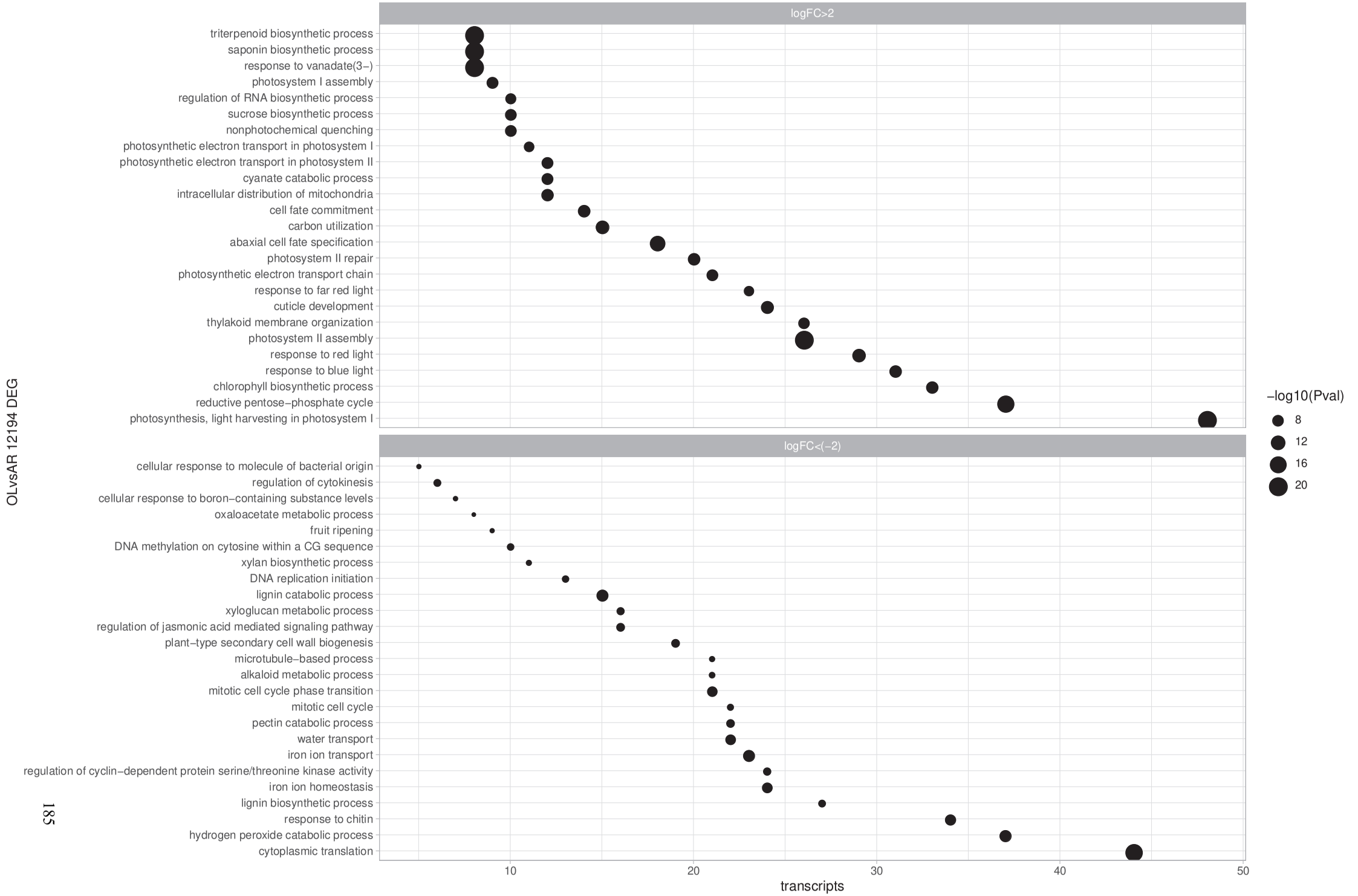


Figure S6. Significantly enriched GO biological pathways (p<0.05) in the upregulated (log fold change > 2) and downregulated (log fold change <2) OLvsAR DGE comparison.

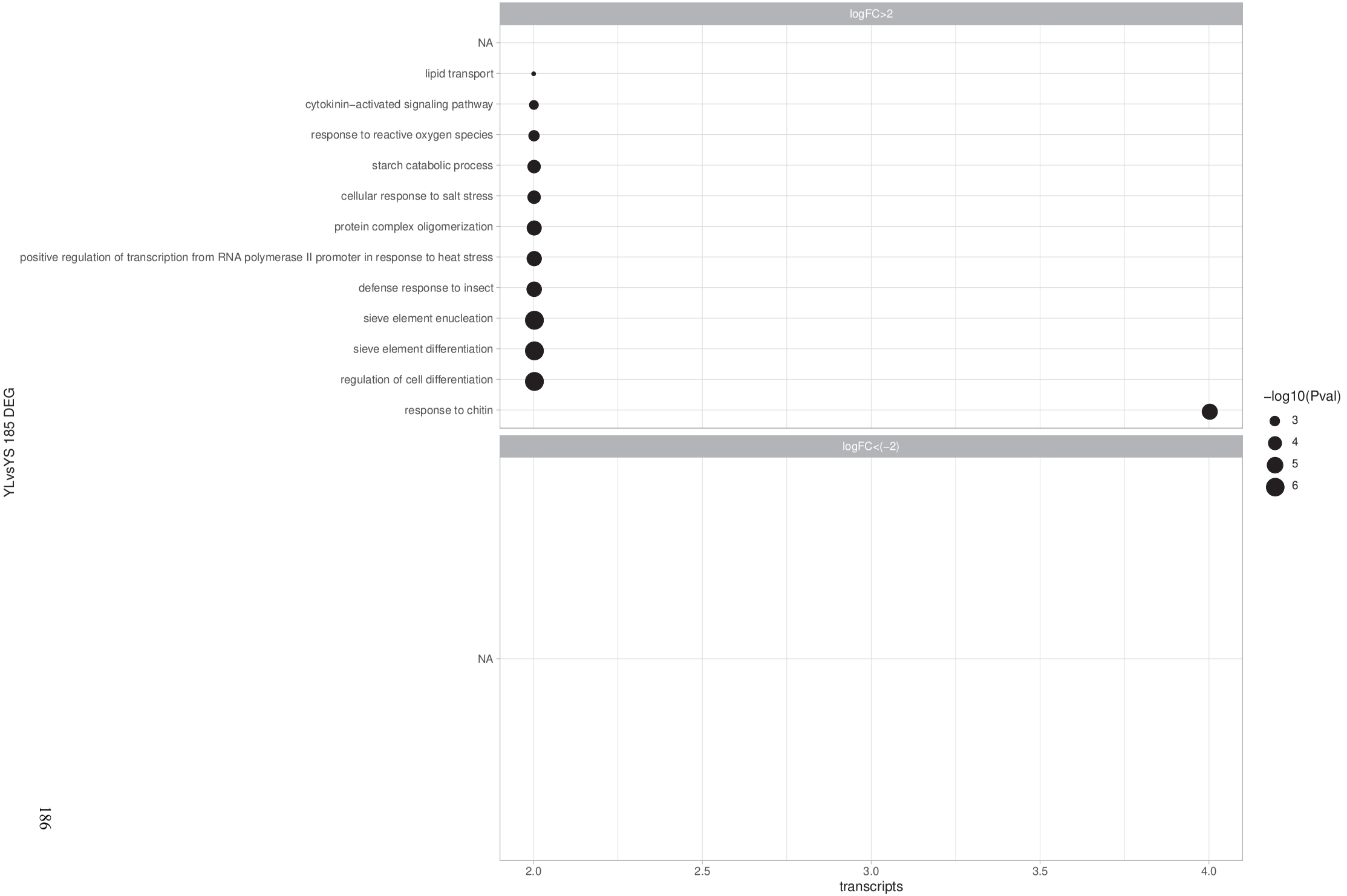


Figure S7. Significantly enriched GO biological pathways (p<0.05) in the upregulated (log fold change > 2) and downregulated (log fold change <2) YLvsYS DEG comparison.

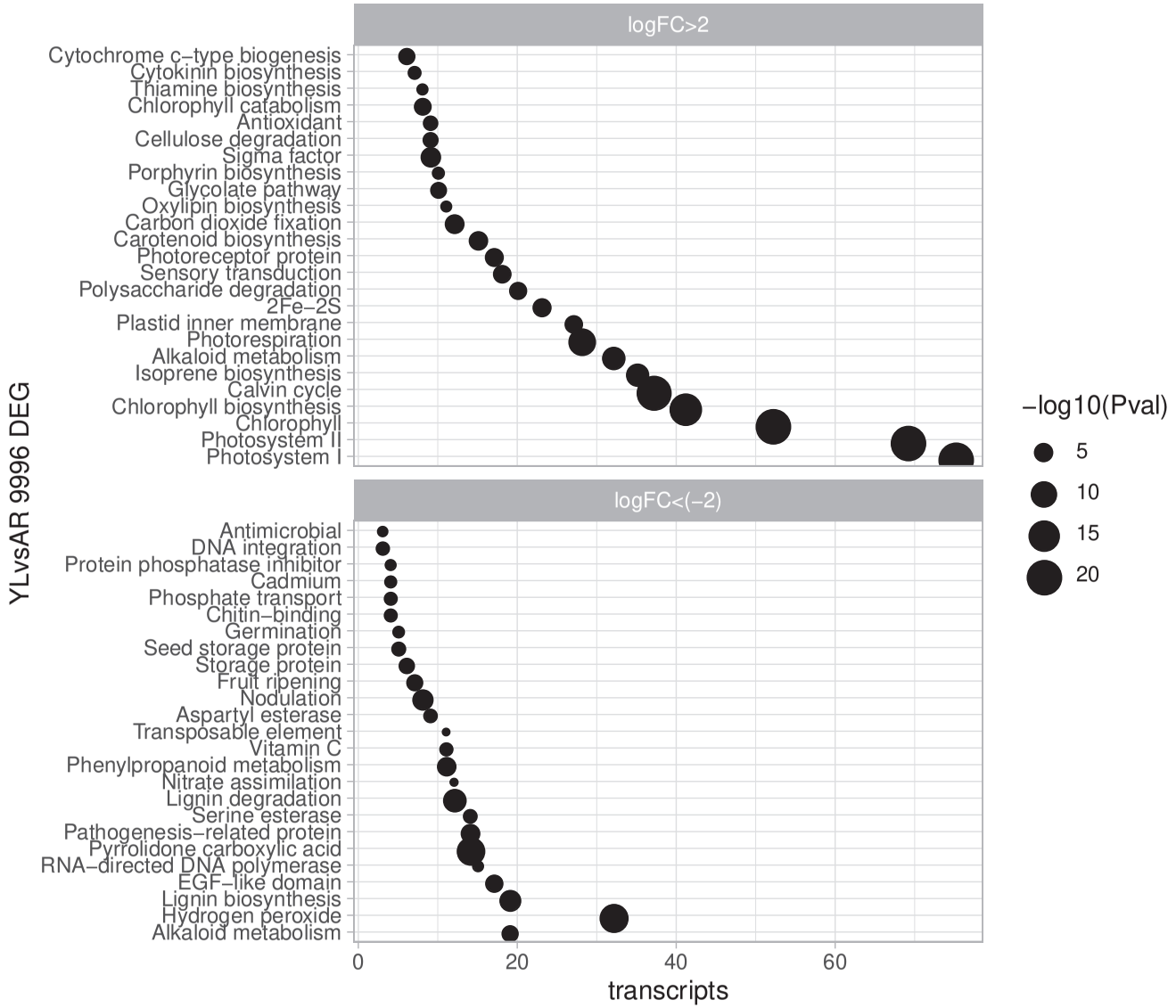


Figure S8. Significantly enriched GO biological pathways (p<0.05) in the upregulated (log fold change > 2) and downregulated (log fold change <2) YLvsAR

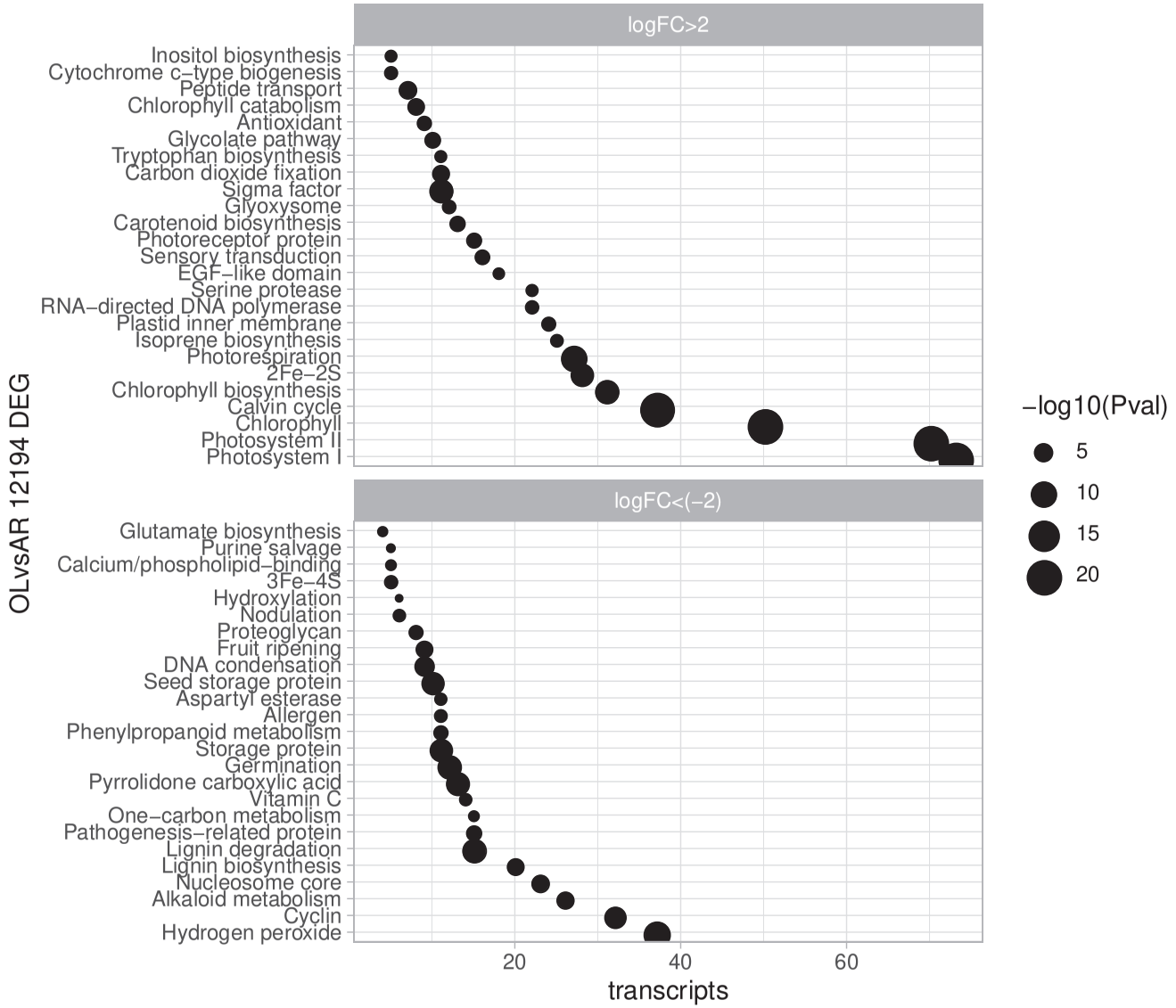


Figure S9. Significantly enriched GO Keywords (p<0.05) in the upregulated (log fold change > 2) and downregulated (log fold change >2) OLvsAR DEG comparison.

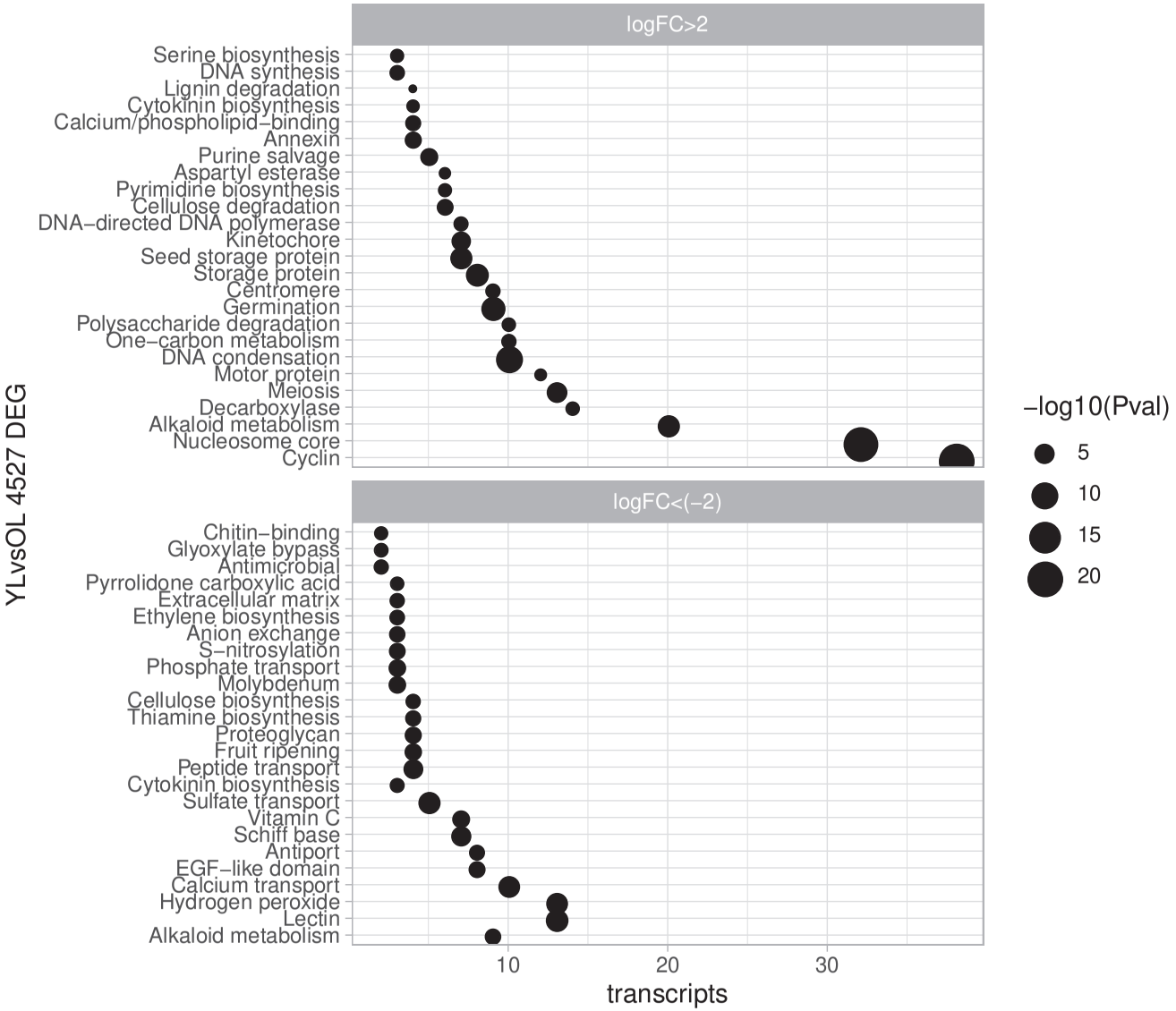


Figure S10. Significantly enriched GO Keywords ($p<0.05$) in the upregulated (log fold change > 2) and downregulated (log fold change <2) YLvsOL DEG comparison.

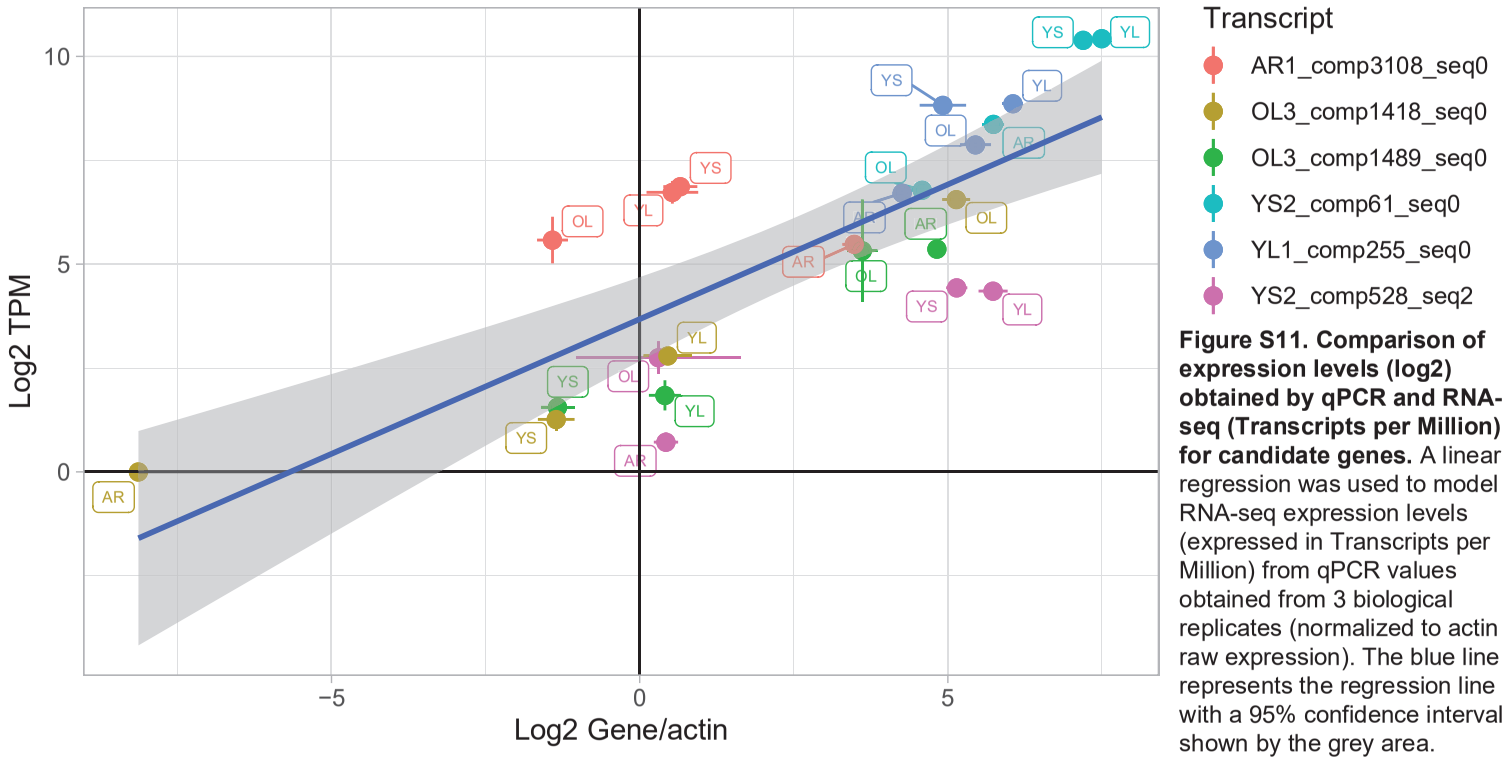


Figure S11. Comparison of expression levels (log2) obtained by qPCR and RNA-seq (Transcripts per Million) for candidate genes. A linear regression was used to model RNA-seq expression levels (expressed in Transcripts per Million) from qPCR values obtained from 3 biological replicates (normalized to actin raw expression). The blue line represents the regression line with a 95% confidence interval shown by the grey area.

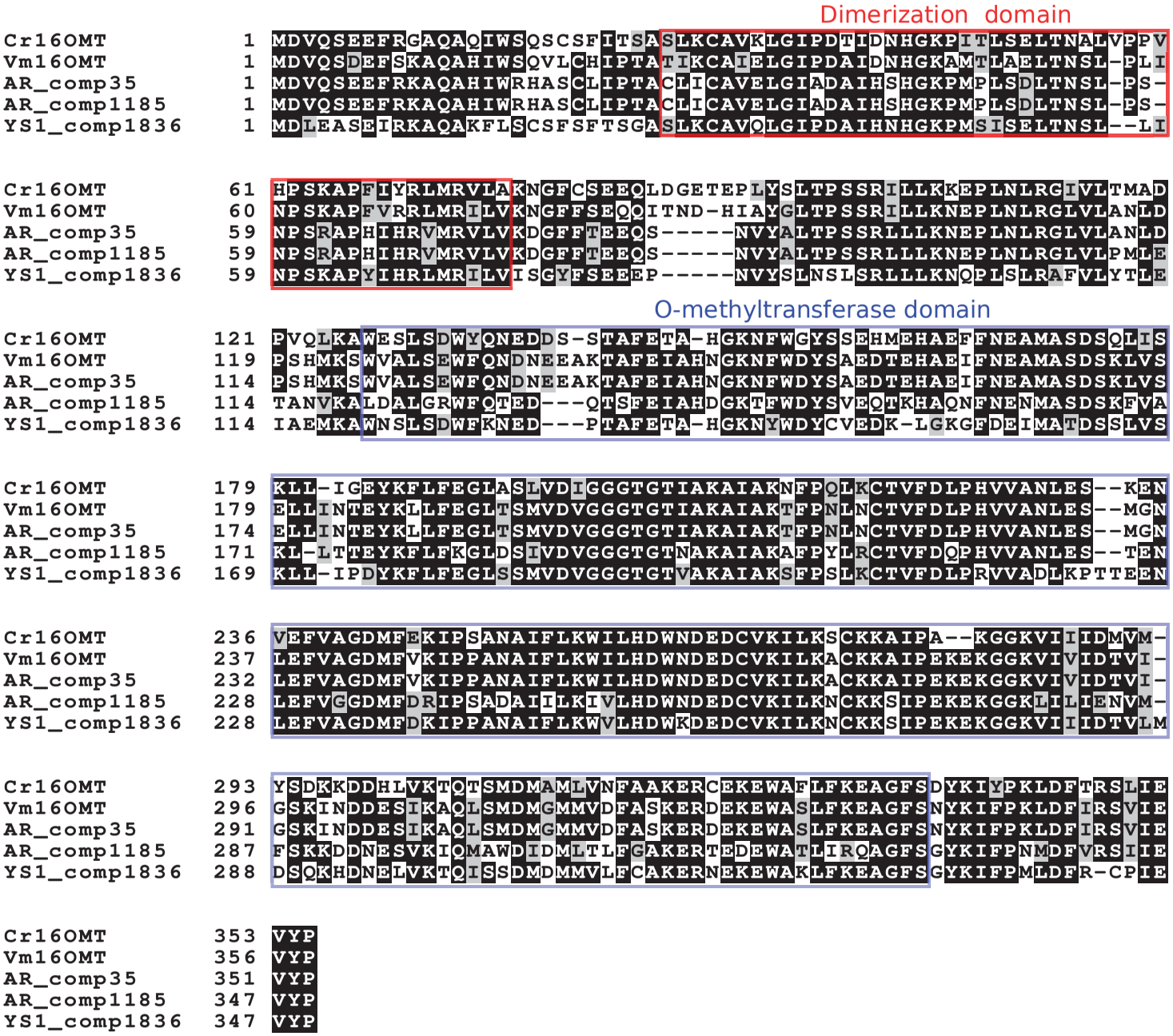


Figure S12. Alignment of the amino-acid sequences of 16-O-tabersonine methyltransferase candidates. Sequence identity and similarity are highlighted by black and grey shading, respectively. The dimerization domain is highlighted in red and the O-methyltransferase domain is highlighted in blue.

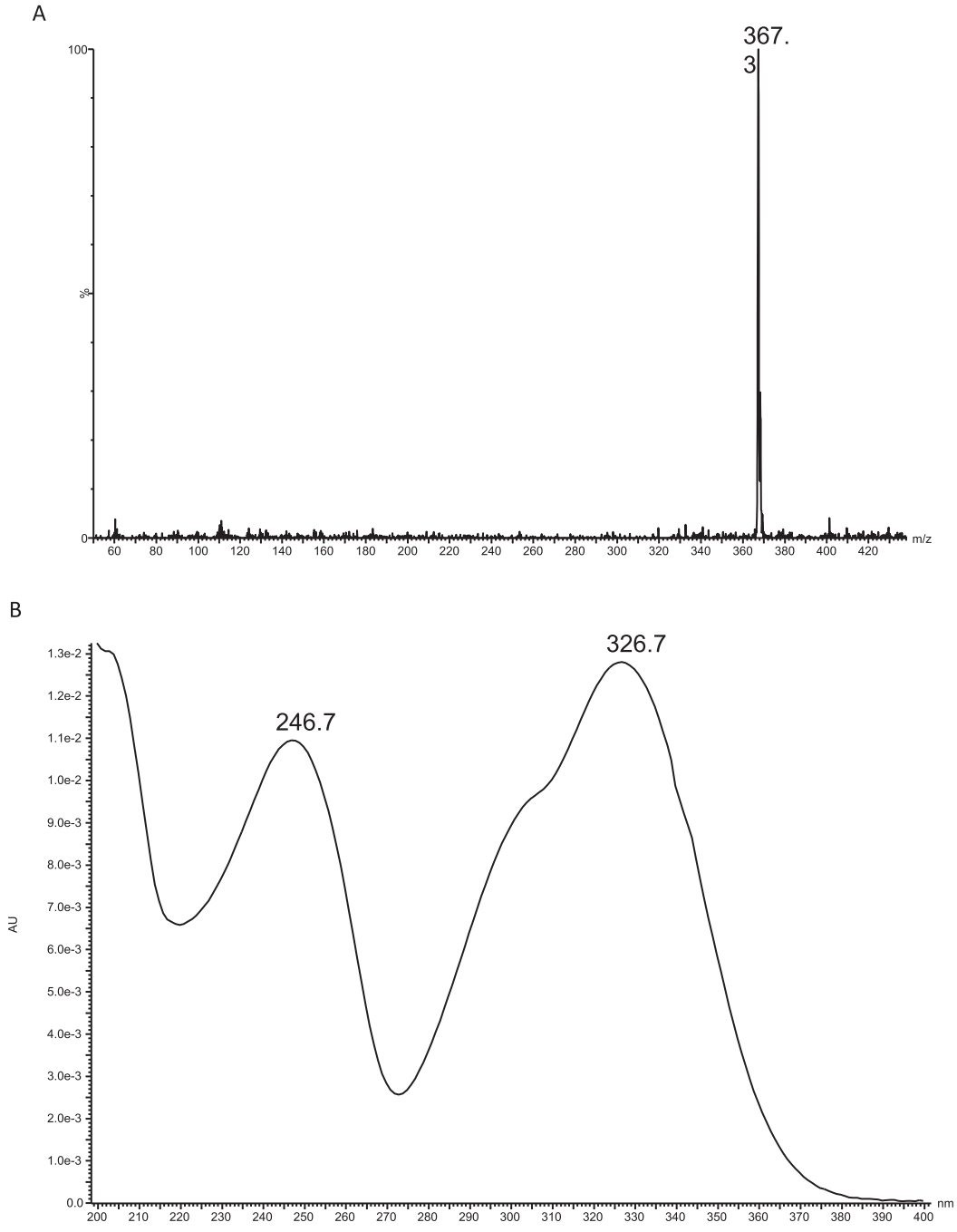


Figure S13. UPLC-DAD-MS analysis of 16-methoxytabersonine. (A) MS spectrum and (B) UV spectrum

TableS1

MIA	Transcript	% ID	Alignment length	Mismatch	Gapopen	MIA start	MIA end	Transcript start	Transcript end	E-value	Bitscore
	YS2_GAGATTCC-ATAGAGGC_comp2127_seq1	85.856	2326	280	9	132	2453	273	2553	0	2715
	YL3_CGCTCATT-ATAGAGGC_comp1779_seq8	86.735	2141	244	7	132	2268	247	2351	0	2585
	YL3_CGCTCATT-ATAGAGGC_comp1779_seq8	81.771	192	26	2	2262	2453	2431	2613	2.94E-46	188
	YL3_CGCTCATT-ATAGAGGC_comp1779_seq5	86.933	1852	199	9	138	1985	398	2210	0	2248
	YL3_CGCTCATT-ATAGAGGC_comp1779_seq5	85.614	285	41	0	1984	2268	2294	2578	1.33E-88	329
	YL3_CGCTCATT-ATAGAGGC_comp1779_seq5	81.771	192	26	2	2262	2453	2658	2840	2.94E-46	188
	AR1_GAGATTCC-CCTATCCT_comp8926_seq1	81.237	1956	367	0	468	2423	1115	3070	0	1873
	OL3_CGCTCATT-GGCTCTGA_comp1793_seq1	85.478	1267	144	7	132	1394	226	1456	0	1460
	OL3_CGCTCATT-GGCTCTGA_comp1793_seq1	87.184	593	74	2	1393	1985	1593	2183	0	720
DXS1 (KC625536.1)	OL3_CGCTCATT-GGCTCTGA_comp1793_seq1	84.12	466	65	2	1984	2449	2267	2723	8.38E-142	506
	YL1_CGCTCATT-TATAGCCT_comp1491_seq4	73.777	1922	496	4	444	2358	295	2215	0	1184
	AR1_GAGATTCC-CCTATCCT_comp8926_seq1	72.282	1959	536	6	444	2397	1099	3055	0	1065
	YS2_GAGATTCC-ATAGAGGC_comp2127_seq1	71.252	1941	541	14	467	2397	584	2517	0	937
	YS1_GAGATTCC-TATAGCCT_comp16391_seq0	74.211	1489	371	7	878	2358	1717	3200	0	935
	YL1_CGCTCATT-TATAGCCT_comp3571_seq0	84.672	822	126	0	871	1692	1	822	0	915
	YL1_CGCTCATT-TATAGCCT_comp3571_seq0	87.847	288	35	0	1970	2257	2335	2622	2.21E-98	361
	YL1_CGCTCATT-TATAGCCT_comp3571_seq0	87.943	282	34	0	1684	1965	1564	1845	9.42E-97	356
DXS2 (DQ848672.1)	YL1_CGCTCATT-TATAGCCT_comp3571_seq0	86.792	159	21	0	2256	2414	3195	3353	2.37E-47	192
	YL1_CGCTCATT-TATAGCCT_comp1491_seq4	87.967	2169	243	5	32	2185	94	2259	0	2729
	YS1_GAGATTCC-TATAGCCT_comp16391_seq0	88.066	1525	182	0	661	2185	1720	3244	0	1930
	AR1_GAGATTCC-CCTATCCT_comp8926_seq1	77.923	1993	430	4	127	2114	1002	2989	0	1602
	YS2_GAGATTCC-ATAGAGGC_comp2127_seq1	70.629	1859	540	4	259	2114	596	2451	0	879
DXS3 (AJ011840.2)	YS2_GAGATTCC-ATAGAGGC_comp28794_seq0	85.752	751	92	4	34	772	50	797	0	868
	YS1_GAGATTCC-TATAGCCT_comp445_seq2	88.005	1634	157	14	1	1613	352	1967	0	2037
	YL3_CGCTCATT-ATAGAGGC_comp492_seq0	87.636	1561	155	14	96	1629	728	2277	0	1918
	YL3_CGCTCATT-ATAGAGGC_comp492_seq0	90.816	98	9	0	1	98	300	397	1.00E-30	136
	YL1_CGCTCATT-TATAGCCT_comp16369_seq2	87.627	1382	133	14	275	1629	849	2219	0	1694
	YL3_CGCTCATT-ATAGAGGC_comp434_seq2	92.004	1113	86	1	1	1110	213	1325	0	1606
	YL3_CGCTCATT-ATAGAGGC_comp434_seq2	79.89	547	72	14	1110	1629	1472	2007	5.51E-129	462
	YS1_GAGATTCC-TATAGCCT_comp445_seq3	91.982	1110	89	0	1	1110	352	1461	0	1602
	YS1_GAGATTCC-TATAGCCT_comp445_seq3	75.75	400	59	14	1257	1629	1892	2280	1.48E-66	255
DXR (AF250235.1)	YS1_GAGATTCC-TATAGCCT_comp445_seq3	91.156	147	13	0	1110	1256	1620	1766	6.74E-52	206
	YS2_GAGATTCC-ATAGAGGC_comp2915_seq1	84.587	1077	120	13	38	1075	1351	2420	0	1177
	YL1_CGCTCATT-TATAGCCT_comp25159_seq0	74.118	85	10	2	949	1024	1	82	6.35E-07	57.2
	YS2_GAGATTCC-ATAGAGGC_comp22987_seq0	95.652	23	1	0	384	406	1	23	0.6	37.4
	OL1_CGCTCATT-CCTATCCT_comp6333_seq2	89.286	28	3	0	662	689	477	450	0.6	37.4
CMS (FJ177510.1)	AR2_GAGATTCC-GGCTCTGA_comp22244_seq0	92.593	27	0	1	1336	1362	25	1	0.6	37.4
	YS1_GAGATTCC-TATAGCCT_comp1496_seq2	84.572	1426	186	12	90	1504	245	1647	0	1557
	AR1_GAGATTCC-CCTATCCT_comp2159_seq1	84.561	1425	187	12	90	1504	281	1682	0	1553
	YL3_CGCTCATT-ATAGAGGC_comp1768_seq0	88.545	969	111	0	90	1058	234	1202	0	1247
	YL3_CGCTCATT-ATAGAGGC_comp1768_seq0	88.136	59	7	0	1227	1285	1719	1777	2.99E-12	75.2

TableS1

	YL3_CGCTCATT-ATAGAGGC_comp1768_seq0	85.417	48	7	0	1169	1216	1199	1246	2.80E-06	55.4
	AR1_GAGATTCC-CCTATCCT_comp2159_seq2	92.323	495	38	0	638	1132	941	1435	0	722
	AR1_GAGATTCC-CCTATCCT_comp2159_seq2	85.278	557	82	0	90	646	281	837	0	636
	AR1_GAGATTCC-CCTATCCT_comp2159_seq2	73.438	384	69	12	1131	1504	1675	2035	2.01E-52	208
CMK (DQ848671.1)	YL3_CGCTCATT-ATAGAGGC_comp18_seq9	76.907	485	78	12	1031	1504	1869	2330	3.18E-94	347
	YS1_GAGATTCC-TATAGCCT_comp2048_seq1	85.405	740	78	7	1	719	179	909	0	843
	AR2_GAGATTCC-GGCTCTGA_comp29267_seq0	92.722	371	27	0	346	716	1002	1372	9.27E-155	547
	YL3_CGCTCATT-ATAGAGGC_comp2559_seq2	82.318	509	60	7	1	488	87	586	8.13E-143	508
	YL3_CGCTCATT-ATAGAGGC_comp2559_seq2	90.431	209	20	0	478	686	1498	1706	1.38E-76	288
	AR2_GAGATTCC-GGCTCTGA_comp5554_seq2	85.676	377	42	3	124	488	197	573	1.47E-120	434
	AR2_GAGATTCC-GGCTCTGA_comp5554_seq2	92.975	242	17	0	478	719	1539	1780	2.66E-98	360
	AR2_GAGATTCC-GGCTCTGA_comp5554_seq2	87.805	41	5	0	1	41	89	129	1.80E-05	51.8
MECS (AF250236.1)	AR1_GAGATTCC-CCTATCCT_comp4429_seq6	81.579	38	7	0	845	882	202	165	0.4	37.4
	YS2_GAGATTCC-ATAGAGGC_comp287_seq1	89.876	2331	220	6	42	2364	139	2461	0	3128
	AR1_GAGATTCC-CCTATCCT_comp972_seq1	89.79	2331	222	6	42	2364	135	2457	0	3119
	AR1_GAGATTCC-CCTATCCT_comp972_seq2	89.79	2331	222	6	42	2364	135	2457	0	3119
	AR1_GAGATTCC-CCTATCCT_comp972_seq6	89.79	2331	222	6	42	2364	135	2457	0	3119
	OL1_CGCTCATT-CCTATCCT_comp301_seq2	90.707	2249	209	0	116	2364	398	2646	0	3113
HDS (AY184810.2)	OL1_CGCTCATT-CCTATCCT_comp301_seq2	92.308	26	2	0	52	77	206	231	0.3	39.2
	YS1_GAGATTCC-TATAGCCT_comp439_seq2	88.865	1419	122	3	1	1392	147	1556	0	1865
	AR2_GAGATTCC-GGCTCTGA_comp1917_seq2	89.177	1349	113	2	1	1325	88	1427	0	1795
	AR2_GAGATTCC-GGCTCTGA_comp1917_seq2	86.486	74	7	1	1322	1392	1580	1653	3.82E-16	87.8
	YS1_GAGATTCC-TATAGCCT_comp439_seq0	92.056	1133	87	1	263	1392	882	2014	0	1636
	YS1_GAGATTCC-TATAGCCT_comp439_seq0	79.508	244	33	5	40	267	3	245	1.62E-52	208
	OL3_CGCTCATT-GGCTCTGA_comp26094_seq0	91.56	936	76	1	460	1392	934	1869	0	1330
	OL3_CGCTCATT-GGCTCTGA_comp26094_seq0	93.939	198	12	0	265	462	3	200	2.74E-81	304
	YL3_CGCTCATT-ATAGAGGC_comp21618_seq1	91.799	756	59	1	640	1392	410	1165	0	1083
HDR (DQ848676.1)	YL3_CGCTCATT-ATAGAGGC_comp21618_seq1	93.145	248	17	0	397	644	1	248	2.25E-101	370
	AR1_GAGATTCC-CCTATCCT_comp1158_seq0	81.466	1187	145	25	36	1206	120	1247	0	1104
	YS2_GAGATTCC-ATAGAGGC_comp566_seq2	84.258	775	89	10	388	1156	1384	2131	0	832
	YS2_GAGATTCC-ATAGAGGC_comp566_seq2	77.344	256	28	7	137	389	19	247	9.33E-49	196
	AR2_GAGATTCC-GGCTCTGA_comp16125_seq3	83.484	775	95	10	388	1156	3012	3759	0	805
	AR2_GAGATTCC-GGCTCTGA_comp16125_seq3	79.31	58	1	1	332	389	1829	1875	5.90E-07	57.2
	YS2_GAGATTCC-ATAGAGGC_comp29331_seq0	92.727	165	12	0	315	479	1	165	2.05E-63	244
	YS2_GAGATTCC-ATAGAGGC_comp29331_seq0	92.391	92	7	0	477	568	1003	1094	2.68E-30	134
IDI1 (EU135981.1)	YS1_GAGATTCC-TATAGCCT_comp15918_seq0	96.552	58	2	0	332	389	1203	1260	6.72E-19	96.9
	AR1_GAGATTCC-CCTATCCT_comp8039_seq0	87.905	1265	147	2	2	1263	301	1562	0	1588
	AR1_GAGATTCC-CCTATCCT_comp30038_seq0	89.655	29	3	0	1196	1224	30	2	0.16	39.2
	YL3_CGCTCATT-ATAGAGGC_comp1244_seq0	95.455	22	1	0	39	60	3610	3589	1.9	35.6
	AR2_GAGATTCC-GGCTCTGA_comp5097_seq2	91.667	24	2	0	691	714	2295	2318	1.9	35.6
GPPS (KF561462.1)	YL3_CGCTCATT-ATAGAGGC_comp1244_seq6	95.455	22	1	0	39	60	3498	3477	1.9	35.6

TableS1

	YL1_CGCTCATT-TATAGCCT_comp695_seq0	85.882	1615	215	6	153	1762	323	1929	0	1869
	YL1_CGCTCATT-TATAGCCT_comp695_seq0	89.189	37	4	0	91	127	285	321	1.23E-04	50
	YS1_GAGATTCC-TATAGCCT_comp841_seq0	87.888	677	77	2	1090	1762	1649	2324	0	847
	OL3_CGCTCATT-GGCTCTGA_comp6736_seq2	80.46	783	121	5	93	874	185	936	0	729
	YL3_CGCTCATT-ATAGAGGC_comp13409_seq0	86.965	537	65	2	1230	1762	874	1409	0	648
	YL3_CGCTCATT-ATAGAGGC_comp13409_seq0	95.833	24	1	0	1090	1113	390	413	0.22	39.2
	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	87.847	288	35	0	943	1230	2	289	1.49E-98	361
	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	88.764	89	10	0	1481	1569	1169	1257	1.01E-24	116
GES (JN882024.1)	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	95.833	24	1	0	1230	1253	726	749	0.22	39.2
	AR1_GAGATTCC-CCTATCCT_comp589_seq7	79.961	1537	300	2	23	1551	1306	2842	0	1382
	AR1_GAGATTCC-CCTATCCT_comp2452_seq0	81.882	861	156	0	316	1176	1	861	0	850
	AR1_GAGATTCC-CCTATCCT_comp2452_seq0	79.191	173	26	2	1385	1555	866	1030	4.58E-35	150
	YL3_CGCTCATT-ATAGAGGC_comp1775_seq0	80.248	805	149	2	753	1555	2	798	0	735
	YS1_GAGATTCC-TATAGCCT_comp8154_seq1	80.863	371	61	2	1187	1555	5	367	8.82E-95	349
G10H (AJ251269.1)	AR1_GAGATTCC-CCTATCCT_comp7341_seq0	73.913	184	48	0	1164	1347	1434	1617	9.45E-25	116
	YL3_CGCTCATT-ATAGAGGC_comp70_seq0	87.871	1113	135	0	129	1241	146	1258	0	1398
	OL3_CGCTCATT-GGCTCTGA_comp191_seq2	86.667	600	80	0	642	1241	1622	2221	0	722
	AR1_GAGATTCC-CCTATCCT_comp635_seq0	68.704	949	280	9	298	1233	2061	3005	2.28E-94	347
	AR1_GAGATTCC-CCTATCCT_comp635_seq0	72.185	151	36	2	136	283	1792	1939	5.04E-14	80.6
	YS1_GAGATTCC-TATAGCCT_comp6458_seq0	82.158	241	43	0	732	972	1	241	2.44E-62	241
10HGO (KF302069.1)	YS2_GAGATTCC-ATAGAGGC_comp2835_seq1	82.553	235	41	0	738	972	1	235	8.52E-62	239
	YL3_CGCTCATT-ATAGAGGC_comp477_seq1	84.9	1053	159	0	115	1167	409	1461	0	1182
	YL3_CGCTCATT-ATAGAGGC_comp477_seq1	86.667	120	16	0	1	120	199	318	1.36E-33	145
	YS1_GAGATTCC-TATAGCCT_comp296_seq0	69.685	1049	292	11	78	1106	302	1344	6.55E-120	432
	OL1_CGCTCATT-CCTATCCT_comp26_seq0	65.448	1094	340	15	82	1153	195	1272	9.75E-61	235
	OL1_CGCTCATT-CCTATCCT_comp11291_seq2	66.289	353	97	9	758	1099	1363	1704	4.74E-14	80.6
	OL1_CGCTCATT-CCTATCCT_comp11291_seq2	95.652	23	0	1	871	893	1552	1573	6.2	33.7
IS (JX974564.1)	AR2_GAGATTCC-GGCTCTGA_comp6919_seq0	84.211	38	5	1	37	74	234	270	0.15	39.2
	OL1_CGCTCATT-CCTATCCT_comp520_seq1	83.735	1537	246	2	14	1548	261	1795	0	1640
	AR2_GAGATTCC-GGCTCTGA_comp3144_seq0	85.755	351	50	0	1198	1548	431	781	9.97E-113	408
	OL3_CGCTCATT-GGCTCTGA_comp13994_seq0	89.416	274	29	0	661	934	3	276	3.72E-99	363
	YS1_GAGATTCC-TATAGCCT_comp11762_seq0	89.286	252	27	0	683	934	1	252	6.30E-90	333
	YS1_GAGATTCC-TATAGCCT_comp11762_seq0	80.682	88	17	0	1206	1293	1686	1773	1.81E-14	82.4
	YS1_GAGATTCC-TATAGCCT_comp17871_seq1	90.244	205	20	0	730	934	1	205	3.49E-74	280
IO (KF302066.1)	YS1_GAGATTCC-TATAGCCT_comp17871_seq1	86.869	99	13	0	933	1031	1040	1138	7.21E-26	120
	YS2_GAGATTCC-ATAGAGGC_comp558_seq3	84.855	1446	219	0	75	1520	832	2277	0	1620
	OL3_CGCTCATT-GGCTCTGA_comp2126_seq0	84.165	1162	184	0	70	1231	72	1233	0	1267
	YS2_GAGATTCC-ATAGAGGC_comp8577_seq0	67.845	1415	415	20	98	1492	162	1556	4.50E-124	446
	YL3_CGCTCATT-ATAGAGGC_comp1412_seq1	83.74	369	60	0	1160	1528	1388	1756	7.13E-109	396
7DLGT (AB733667.1)	YS1_GAGATTCC-TATAGCCT_comp607_seq1	84.928	345	52	0	1184	1528	1063	1407	1.06E-106	389
	YL1_CGCTCATT-TATAGCCT_comp444_seq0	85.597	1222	156	7	662	1874	835	2045	0	1397

TableS1

	YL1_CGCTCATT-TATAGCCT_comp444_seq0	86.364	506	66	1	165	667	236	741	3.58E-170	600
	OL1_CGCTCATT-CCTATCCT_comp2662_seq1	68.421	1121	329	13	614	1726	783	1886	3.36E-107	390
	OL1_CGCTCATT-CCTATCCT_comp2662_seq1	73.469	98	26	0	233	330	408	505	7.44E-08	60.8
	AR2_GAGATTCC-GGCTCTGA_comp35_seq9	66.021	1498	477	14	233	1717	742	2220	9.02E-102	372
	OL3_CGCTCATT-GGCTCTGA_comp1360_seq3	68.489	1006	280	16	735	1726	1112	2094	3.59E-94	347
7DLH (KF415115.1)	AR2_GAGATTCC-GGCTCTGA_comp17345_seq0	67.033	1183	349	16	569	1735	494	1651	5.33E-92	340
	YS2_GAGATTCC-ATAGAGGC_comp88_seq3	85.714	1078	151	1	39	1116	385	1459	0	1249
	YS2_GAGATTCC-ATAGAGGC_comp1328_seq0	86.269	1005	135	1	112	1116	5289	6290	0	1189
	YL1_CGCTCATT-TATAGCCT_comp5423_seq0	84.903	828	122	1	39	866	2873	3697	0	928
	YL1_CGCTCATT-TATAGCCT_comp5423_seq0	89.535	344	36	0	773	1116	2120	2463	4.50E-128	459
	YL1_CGCTCATT-TATAGCCT_comp5423_seq0	83.871	31	5	0	551	581	878	908	5.9	33.7
	OL3_CGCTCATT-GGCTCTGA_comp1728_seq2	83.862	756	116	2	24	779	238	987	0	810
	OL3_CGCTCATT-GGCTCTGA_comp1728_seq2	88.663	344	39	0	773	1116	2169	2512	9.91E-124	444
LAMT (EU057974.1)	YS1_GAGATTCC-TATAGCCT_comp18495_seq0	88.53	558	64	0	559	1116	2842	3399	0	719
	AR2_GAGATTCC-GGCTCTGA_comp35_seq9	86.951	1617	203	2	15	1630	631	2240	0	1963
	YS1_GAGATTCC-TATAGCCT_comp24972_seq0	87.259	1295	165	0	342	1636	1436	2730	0	1591
	YS1_GAGATTCC-TATAGCCT_comp24972_seq0	86.722	241	32	0	105	345	3	243	2.32E-77	291
	YS1_GAGATTCC-TATAGCCT_comp24972_seq0	72.973	74	5	3	1661	1732	2785	2845	0.005	44.6
	YL1_CGCTCATT-TATAGCCT_comp7664_seq0	86.38	1138	155	0	499	1636	317	1454	0	1353
	YL1_CGCTCATT-TATAGCCT_comp7664_seq0	83.929	168	27	0	63	230	150	317	3.02E-44	181
	YL1_CGCTCATT-TATAGCCT_comp7664_seq0	72.973	74	5	3	1661	1732	1509	1569	0.005	44.6
	YL1_CGCTCATT-TATAGCCT_comp14816_seq0	86.883	831	109	0	806	1636	424	1254	0	1007
	YL1_CGCTCATT-TATAGCCT_comp14816_seq0	89.6	250	26	0	565	814	1	250	2.17E-90	334
	YL1_CGCTCATT-TATAGCCT_comp14816_seq0	72.973	74	5	3	1661	1732	1309	1369	0.005	44.6
	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	86.845	821	108	0	816	1636	478	1298	0	994
	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	86.264	182	25	0	52	233	101	282	1.46E-54	215
SLS1 (L10081.1)	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	72.973	74	5	3	1661	1732	1353	1413	0.005	44.6
	AR2_GAGATTCC-GGCTCTGA_comp35_seq9	87.626	1584	196	0	1	1584	667	2250	0	1974
	YS1_GAGATTCC-TATAGCCT_comp24972_seq0	87.792	1286	157	0	286	1571	1436	2721	0	1611
	YS1_GAGATTCC-TATAGCCT_comp24972_seq0	86.307	241	33	0	49	289	3	243	8.40E-76	286
	YL1_CGCTCATT-TATAGCCT_comp7664_seq0	86.714	1129	150	0	443	1571	317	1445	0	1361
	YL1_CGCTCATT-TATAGCCT_comp7664_seq0	82.738	168	29	0	7	174	150	317	1.33E-41	172
	YL1_CGCTCATT-TATAGCCT_comp14816_seq0	87.698	821	101	0	751	1571	425	1245	0	1025
	YL1_CGCTCATT-TATAGCCT_comp14816_seq0	86.8	250	33	0	509	758	1	250	1.09E-80	302
	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	87.562	812	101	0	760	1571	478	1289	0	1009
	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	84.746	177	27	0	1	177	106	282	3.35E-49	197
SLS2 (KF309242.1)	YS2_GAGATTCC-ATAGAGGC_comp2723_seq0	90	30	3	0	503	532	454	483	0.057	41
	known MIA genes from C. roseu	78.534	955	195	4	44	991	102	1053	0	789
	YS2_GAGATTCC-ATAGAGGC_comp2161_seq0	82.051	39	7	0	815	853	8327	8365	0.15	39.2
	YL3_CGCTCATT-ATAGAGGC_comp4991_seq1	82.051	39	7	0	815	853	2791	2829	0.15	39.2
	OL1_CGCTCATT-CCTATCCT_comp1693_seq1	92.308	26	2	0	904	929	1163	1188	0.15	39.2

STR (X53602.1)	YL3_CGCTCATT-ATAGAGGC_comp10885_seq1	95.652	23	1	0	1045	1067	1410	1388	0.51	37.4
	YS2_GAGATTCC-ATAGAGGC_comp324_seq0	80.227	1057	201	4	1	1050	661	1716	0	953
	YS2_GAGATTCC-ATAGAGGC_comp324_seq0	75.844	385	88	2	1050	1433	1786	2166	1.67E-71	271
	YL3_CGCTCATT-ATAGAGGC_comp4131_seq0	70.44	159	45	2	824	981	1727	1884	3.69E-10	68
	YS1_GAGATTCC-TATAGCCT_comp9626_seq1	84.211	38	5	1	831	868	668	632	0.18	39.2
	AR1_GAGATTCC-CCTATCCT_comp8861_seq1	100	21	0	0	437	457	1687	1667	0.18	39.2
TDC (MG748691.1)	YL3_CGCTCATT-ATAGAGGC_comp9172_seq1	84.211	38	5	1	831	868	659	623	0.18	39.2
	YL1_CGCTCATT-TATAGCCT_comp255_seq0	79.31	1537	265	11	133	1662	1595	3085	0	1337
	OL1_CGCTCATT-CCTATCCT_comp1264_seq0	77.739	867	184	5	417	1277	391	1254	0	679
	OL1_CGCTCATT-CCTATCCT_comp1264_seq0	85.256	156	23	0	1430	1585	1251	1406	3.30E-43	178
	OL1_CGCTCATT-CCTATCCT_comp1264_seq0	86.429	140	19	0	263	402	76	215	5.97E-40	167
	OL3_CGCTCATT-GGCTCTGA_comp686_seq1	83.443	610	98	2	478	1084	958	1567	0	639
	OL3_CGCTCATT-GGCTCTGA_comp686_seq1	87.931	348	40	2	133	479	517	863	9.43E-120	432
	OL3_CGCTCATT-GGCTCTGA_comp686_seq1	73.469	490	100	3	1173	1662	1838	2297	2.22E-83	311
	YL3_CGCTCATT-ATAGAGGC_comp1113_seq0	71.298	655	152	8	931	1580	2	625	2.37E-89	331
	AR2_GAGATTCC-GGCTCTGA_comp5833_seq1	67.018	949	282	8	142	1082	378	1303	5.96E-78	293
SGD (EU072423.1)	AR2_GAGATTCC-GGCTCTGA_comp5833_seq1	67.974	153	46	1	1417	1566	1608	1760	9.48E-06	53.6
	OL1_CGCTCATT-CCTATCCT_comp374_seq0	86.575	1095	147	0	1	1095	176	1270	0	1312
	AR2_GAGATTCC-GGCTCTGA_comp706_seq1	86.066	1098	150	1	1	1095	61	1158	0	1288
	AR2_GAGATTCC-GGCTCTGA_comp706_seq3	86.066	1098	150	1	1	1095	61	1158	0	1288
	AR2_GAGATTCC-GGCTCTGA_comp706_seq4	86.066	1098	150	1	1	1095	61	1158	0	1288
GS1 (KF302079.1)	YL3_CGCTCATT-ATAGAGGC_comp123_seq0	85.845	1095	155	0	1	1095	2080	3174	0	1276
	OL1_CGCTCATT-CCTATCCT_comp374_seq0	84.081	1093	168	1	1	1087	176	1268	0	1187
	YL3_CGCTCATT-ATAGAGGC_comp123_seq0	83.715	1093	172	1	1	1087	2080	3172	0	1169
	AR2_GAGATTCC-GGCTCTGA_comp706_seq1	83.212	1096	175	2	1	1087	61	1156	0	1146
	AR2_GAGATTCC-GGCTCTGA_comp706_seq3	83.212	1096	175	2	1	1087	61	1156	0	1146
GS2 (KF302078.1)	AR2_GAGATTCC-GGCTCTGA_comp706_seq4	83.212	1096	175	2	1	1087	61	1156	0	1146
	AR1_GAGATTCC-CCTATCCT_comp3108_seq0	81.699	1071	181	2	6	1064	157	1224	0	1052
	OL1_CGCTCATT-CCTATCCT_comp14114_seq0	74.637	757	160	7	325	1062	501	1244	1.72E-139	497
	YL3_CGCTCATT-ATAGAGGC_comp42_seq0	69.436	1047	276	11	40	1052	177	1213	4.92E-121	435
	AR1_GAGATTCC-CCTATCCT_comp80_seq0	67.739	1057	305	10	38	1064	287	1337	3.79E-97	356
THAS1 (KM524258.1)	AR1_GAGATTCC-CCTATCCT_comp3050_seq0	67.739	1057	305	10	38	1064	281	1331	3.79E-97	356
	OL1_CGCTCATT-CCTATCCT_comp3640_seq0	75.909	1100	260	3	18	1116	210	1305	0	780
	YS2_GAGATTCC-ATAGAGGC_comp697_seq4	78.672	783	162	3	11	792	98	876	0	650
	YS2_GAGATTCC-ATAGAGGC_comp697_seq4	72.566	113	31	0	949	1061	1033	1145	3.48E-09	64.4
	OL3_CGCTCATT-GGCTCTGA_comp1793_seq1	73.4	1094	284	5	25	1116	2866	3954	0	645
	YS2_GAGATTCC-ATAGAGGC_comp4098_seq3	73.4	1094	284	5	25	1116	4818	5906	0	645
THAS2 (KU865323.1)	YL1_CGCTCATT-TATAGCCT_comp13612_seq0	75.938	906	213	3	206	1110	847	1748	0	643
	OL1_CGCTCATT-CCTATCCT_comp3640_seq0	81.088	1121	192	6	18	1134	201	1305	0	1058
	OL3_CGCTCATT-GGCTCTGA_comp1793_seq1	80.454	926	161	6	213	1134	3045	3954	0	845
	YS2_GAGATTCC-ATAGAGGC_comp4098_seq3	80.454	926	161	6	213	1134	4997	5906	0	845

TableS1

	YL1_CGCTCATT-TATAGCCT_comp13612_seq0	79.563	915	173	4	215	1128	847	1748	0	801
THAS3 (KU865322.1)	YS2_GAGATTCC-ATAGAGGC_comp697_seq4	70.31	1098	302	9	18	1109	96	1175	7.74E-138	491
	AR1_GAGATTCC-CCTATCCT_comp3108_seq0	81.648	1068	194	2	1	1067	158	1224	0	1036
	OL1_CGCTCATT-CCTATCCT_comp14114_seq0	76.923	767	159	7	305	1065	490	1244	9.57E-162	571
	AR1_GAGATTCC-CCTATCCT_comp80_seq0	68.587	1076	320	9	7	1067	265	1337	5.31E-108	392
	OL1_CGCTCATT-CCTATCCT_comp1134_seq1	68.03	1076	326	9	7	1067	141	1213	7.38E-100	365
THAS4 (KU865324.1)	AR1_GAGATTCC-CCTATCCT_comp3050_seq0	67.844	1076	328	9	7	1067	259	1331	3.82E-97	356
	AR1_GAGATTCC-CCTATCCT_comp3108_seq0	82.303	1068	187	2	1	1067	158	1224	0	1067
	OL1_CGCTCATT-CCTATCCT_comp14114_seq0	77.004	761	159	6	310	1065	495	1244	9.57E-162	571
	AR1_GAGATTCC-CCTATCCT_comp80_seq0	69.274	1074	316	6	7	1067	265	1337	4.96E-121	435
	OL1_CGCTCATT-CCTATCCT_comp1134_seq1	68.715	1074	322	6	7	1067	141	1213	6.90E-113	408
HYS (KU865325.1)	AR1_GAGATTCC-CCTATCCT_comp3050_seq0	68.063	1074	329	6	7	1067	259	1331	4.08E-103	376
	YL1_CGCTCATT-TATAGCCT_comp32_seq1	82.626	1508	256	2	1	1505	244	1748	0	1535
	YL1_CGCTCATT-TATAGCCT_comp32_seq4	82.626	1508	256	2	1	1505	244	1748	0	1535
	YL3_CGCTCATT-ATAGAGGC_comp58_seq1	82.56	1508	257	2	1	1505	201	1705	0	1530
	YS2_GAGATTCC-ATAGAGGC_comp3760_seq0	84.491	619	96	0	887	1505	992	1610	0	684
GO (MF770508.1)	AR2_GAGATTCC-GGCTCTGA_comp139_seq0	67.253	1478	432	21	7	1461	173	1621	6.98E-121	435
	OL3_CGCTCATT-GGCTCTGA_comp6787_seq2	85.747	863	123	0	1	863	37	899	0	1002
	YL1_CGCTCATT-TATAGCCT_comp16622_seq1	85.315	858	126	0	194	1051	1429	2286	0	980
	YS2_GAGATTCC-ATAGAGGC_comp697_seq1	85.315	858	126	0	194	1051	1531	2388	0	980
	YL3_CGCTCATT-ATAGAGGC_comp42_seq0	81.526	996	184	0	56	1051	223	1218	0	967
redox1 (MF770509.1)	AR1_GAGATTCC-CCTATCCT_comp530_seq0	77.333	975	210	4	84	1051	371	1341	0	755
	YL3_CGCTCATT-ATAGAGGC_comp101_seq3	82.613	972	169	0	1	972	133	1104	0	991
	AR2_GAGATTCC-GGCTCTGA_comp687_seq0	80.844	569	109	0	1	569	98	666	6.19E-151	535
	AR2_GAGATTCC-GGCTCTGA_comp687_seq0	85.217	230	34	0	743	972	1832	2061	5.82E-69	262
	AR2_GAGATTCC-GGCTCTGA_comp687_seq0	87.634	186	23	0	564	749	1009	1194	9.85E-60	232
	YL3_CGCTCATT-ATAGAGGC_comp101_seq2	80.316	569	112	0	1	569	133	701	3.91E-147	522
	YL3_CGCTCATT-ATAGAGGC_comp101_seq2	86.308	409	56	0	564	972	1044	1452	2.81E-136	486
	YS1_GAGATTCC-TATAGCCT_comp65_seq1	80.316	569	112	0	1	569	94	662	3.91E-147	522
	YS1_GAGATTCC-TATAGCCT_comp65_seq1	86.308	409	56	0	564	972	1005	1413	2.81E-136	486
	AR2_GAGATTCC-GGCTCTGA_comp12013_seq0	73.893	858	210	7	37	887	199	1049	4.76E-146	518
redox2 (MF770510.1)	AR2_GAGATTCC-GGCTCTGA_comp12013_seq0	91.667	24	2	0	949	972	1078	1101	1.5	35.6
	YL3_CGCTCATT-ATAGAGGC_comp1258_seq0	79.013	1277	242	7	1	1261	163	1429	0	1086
	AR2_GAGATTCC-GGCTCTGA_comp10632_seq0	71.654	254	61	5	122	368	178	427	1.68E-26	122
	AR2_GAGATTCC-GGCTCTGA_comp10632_seq0	70.588	119	33	2	740	857	835	952	8.72E-05	50
	YL1_CGCTCATT-TATAGCCT_comp10318_seq0	63.83	799	238	18	11	782	67	841	4.51E-21	104
	YL1_CGCTCATT-TATAGCCT_comp10318_seq0	71.895	153	43	0	1089	1241	1184	1336	1.48E-14	82.4
	YS2_GAGATTCC-ATAGAGGC_comp4073_seq1	81.034	58	11	0	1048	1105	1174	1231	2.05E-06	55.4
SAT (MF770511.1)	YS1_GAGATTCC-TATAGCCT_comp9325_seq0	93.75	32	2	0	1073	1104	1188	1219	8.72E-05	50
	YS2_GAGATTCC-ATAGAGGC_comp776_seq0	76.106	1582	344	7	16	1579	126	1691	0	1148
	AR2_GAGATTCC-GGCTCTGA_comp14833_seq0	67.846	1583	453	21	11	1568	4	1555	2.75E-145	517

TableS1

	AR2_GAGATTCC-GGCTCTGA_comp21526_seq0	71.014	414	120	0	397	810	539	952	6.49E-52	206
	AR2_GAGATTCC-GGCTCTGA_comp21526_seq0	71.823	181	45	3	143	317	217	397	1.25E-16	89.7
	AR2_GAGATTCC-GGCTCTGA_comp21526_seq0	72.794	136	36	1	1445	1579	1569	1704	2.76E-12	75.2
	AR1_GAGATTCC-CCTATCCT_comp1722_seq0	70.36	361	107	0	432	792	500	860	1.63E-40	168
PAS	YL1_CGCTCATT-TATAGCCT_comp1884_seq0	67.778	450	145	0	252	701	597	1046	2.95E-37	158
	YL1_CGCTCATT-TATAGCCT_comp695_seq0	85.882	1615	215	6	153	1762	323	1929	0	1869
	YL1_CGCTCATT-TATAGCCT_comp695_seq0	89.189	37	4	0	91	127	285	321	0.000123	50
	YS1_GAGATTCC-TATAGCCT_comp841_seq0	87.888	677	77	2	1090	1762	1649	2324	0	847
	OL3_CGCTCATT-GGCTCTGA_comp6736_seq2	80.46	783	121	5	93	874	185	936	0	729
	YL3_CGCTCATT-ATAGAGGC_comp13409_seq0	86.965	537	65	2	1230	1762	874	1409	0	648
	YL3_CGCTCATT-ATAGAGGC_comp13409_seq0	95.833	24	1	0	1090	1113	390	413	0.22	39.2
	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	87.847	288	35	0	943	1230	2	289	1.49E-98	361
	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	88.764	89	10	0	1481	1569	1169	1257	1.01E-24	116
DPAS (KU865331)	YS1_GAGATTCC-TATAGCCT_comp19336_seq1	95.833	24	1	0	1230	1253	726	749	0.22	39.2
	YL3_CGCTCATT-ATAGAGGC_comp166_seq2	74.669	983	216	9	10	983	284	1242	0	639
	AR1_GAGATTCC-CCTATCCT_comp606_seq1	73.333	510	116	4	475	983	381	871	1.59E-82	307
	AR2_GAGATTCC-GGCTCTGA_comp17902_seq0	72.757	301	64	2	684	984	207	489	1.60E-44	181
	YL1_CGCTCATT-TATAGCCT_comp661_seq0	66.612	614	174	7	25	621	149	748	5.58E-44	179
CS (MF770512.1)	AR1_GAGATTCC-CCTATCCT_comp5172_seq0	78.417	139	30	0	845	983	13	151	5.58E-25	116
	YL3_CGCTCATT-ATAGAGGC_comp166_seq2	80.234	941	177	3	25	962	311	1245	0	854
	AR1_GAGATTCC-CCTATCCT_comp606_seq1	80	485	94	1	478	962	393	874	4.42E-121	435
	YL1_CGCTCATT-TATAGCCT_comp661_seq0	67.885	903	238	16	8	881	147	1026	2.79E-79	297
	AR2_GAGATTCC-GGCTCTGA_comp17902_seq0	80.351	285	56	0	678	962	207	491	5.76E-69	262
TS (MF770513.1)	YS1_GAGATTCC-TATAGCCT_comp13751_seq2	66.332	597	145	17	28	604	1364	1924	1.11E-33	145
	YL3_CGCTCATT-ATAGAGGC_comp166_seq2	79.277	941	189	3	25	965	311	1245	0	810
	AR1_GAGATTCC-CCTATCCT_comp606_seq1	78.745	494	105	0	472	965	381	874	1.19E-115	417
	YL1_CGCTCATT-TATAGCCT_comp661_seq0	67.037	901	256	12	8	886	147	1028	3.90E-71	269
	AR2_GAGATTCC-GGCTCTGA_comp17902_seq0	79.649	285	58	0	681	965	207	491	3.00E-66	253
VS1 (MF770514.1)	OL1_CGCTCATT-CCTATCCT_comp1318_seq1	66.534	502	135	7	28	523	84	558	3.20E-34	147
	YL3_CGCTCATT-ATAGAGGC_comp166_seq2	80.293	888	171	2	49	935	352	1236	0	807
	AR1_GAGATTCC-CCTATCCT_comp606_seq1	80.206	485	96	0	451	935	381	865	2.92E-123	443
	AR2_GAGATTCC-GGCTCTGA_comp17902_seq0	80.42	286	56	0	660	945	207	492	1.62E-69	264
	YL1_CGCTCATT-TATAGCCT_comp661_seq0	66.469	844	250	8	71	891	227	1060	1.85E-62	241
	YL1_CGCTCATT-TATAGCCT_comp661_seq0	79.487	39	8	0	8	46	147	185	1.4	35.6
VS2 (MF770515.1)	AR1_GAGATTCC-CCTATCCT_comp5172_seq0	81.119	143	27	0	793	935	3	145	5.66E-31	136
	YS2_GAGATTCC-ATAGAGGC_comp528_seq2	77.958	1538	304	6	1	1513	2107	3634	0	1251
	YL1_CGCTCATT-TATAGCCT_comp104_seq6	77.893	1538	305	6	1	1513	2274	3801	0	1245
	YS1_GAGATTCC-TATAGCCT_comp15820_seq1	74.339	1438	334	8	107	1513	305	1738	0	926
	YS2_GAGATTCC-ATAGAGGC_comp37475_seq2	79.321	619	113	1	910	1513	998	1616	4.44E-155	549
T16H1 (FJ647194.1)	OL3_CGCTCATT-GGCTCTGA_comp1489_seq0	68.279	1406	400	17	133	1512	943	2328	3.65E-137	489
	YS2_GAGATTCC-ATAGAGGC_comp528_seq2	78.335	1574	314	9	3	1559	2077	3640	0	1285

TableS1

	YL1_CGCTCATT-TATAGCCT_comp104_seq6	78.272	1574	315	9	3	1559	2244	3807	0	1281
	YS1_GAGATTCC-TATAGCCT_comp15820_seq1	74.054	1453	366	4	113	1558	295	1743	0	913
	YS2_GAGATTCC-ATAGAGGC_comp37475_seq2	81.28	625	117	0	935	1559	998	1622	3.13E-170	600
T16H2 (JF742645.1)	OL3_CGCTCATT-GGCTCTGA_comp1489_seq0	68.355	1441	410	21	143	1569	937	2345	4.08E-137	489
	AR2_GAGATTCC-GGCTCTGA_comp35_seq0	75.923	1084	225	6	60	1128	583	1645	0	783
	YS2_GAGATTCC-ATAGAGGC_comp61_seq0	75.668	1085	241	9	60	1128	2794	3871	0	747
	AR1_GAGATTCC-CCTATCCT_comp1185_seq0	72.913	1078	256	6	60	1128	581	1631	1.15E-180	634
16OMT (EF444544.1)	YS1_GAGATTCC-TATAGCCT_comp1836_seq0	72.146	1095	257	9	49	1128	126	1187	1.94E-171	603
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	72.628	917	200	10	1	913	3643	2774	1.77E-147	524
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	82.008	239	43	0	913	1151	2478	2240	3.66E-61	237
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	90	110	11	0	1325	1434	2237	2128	1.46E-34	149
	YL3_CGCTCATT-ATAGAGGC_comp5333_seq0	75.795	723	163	4	192	913	3124	2413	1.36E-142	508
	YL3_CGCTCATT-ATAGAGGC_comp5333_seq0	79.949	394	63	4	1078	1464	2062	1678	2.31E-95	351
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	74.412	723	158	5	192	913	2711	2015	9.79E-132	471
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	86.765	204	20	1	1325	1521	1588	1385	2.02E-64	248
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	79.73	74	15	0	1078	1151	1664	1591	1.37E-09	66.2
	AR1_GAGATTCC-CCTATCCT_comp629_seq5	67.134	1354	379	26	168	1488	192	1512	1.66E-103	378
T3O (LN831958.1)	AR1_GAGATTCC-CCTATCCT_comp5822_seq1	65.632	1385	433	16	112	1485	237	1589	1.36E-85	318
	AR1_GAGATTCC-CCTATCCT_comp3108_seq0	77.788	1049	226	6	22	1068	189	1232	0	821
	OL1_CGCTCATT-CCTATCCT_comp14114_seq0	74.009	681	158	6	396	1071	591	1257	1.37E-116	421
	AR1_GAGATTCC-CCTATCCT_comp80_seq0	67.191	1018	309	12	67	1067	335	1344	1.13E-79	298
	OL1_CGCTCATT-CCTATCCT_comp1134_seq1	66.437	1016	320	10	67	1067	211	1220	5.48E-71	269
	YL3_CGCTCATT-ATAGAGGC_comp42_seq0	78.71	310	63	1	7	313	160	469	2.84E-68	260
T3R (KP122966.1)	YL3_CGCTCATT-ATAGAGGC_comp42_seq0	66.914	538	170	6	515	1048	680	1213	1.29E-34	149
	YL1_CGCTCATT-TATAGCCT_comp302_seq2	77.549	971	192	15	25	978	211	1172	0	724
	YL1_CGCTCATT-TATAGCCT_comp302_seq2	75.8	781	178	7	25	798	1692	2468	7.06E-151	535
	AR2_GAGATTCC-GGCTCTGA_comp630_seq4	76.369	986	207	15	10	978	1276	2252	0	684
	YS1_GAGATTCC-TATAGCCT_comp198_seq4	75.628	796	183	7	10	798	2351	3142	1.66E-152	540
	OL3_CGCTCATT-GGCTCTGA_comp6929_seq0	73.598	856	217	6	20	869	213	1065	3.43E-142	506
NMT (HM584929.1)	YL1_CGCTCATT-TATAGCCT_comp938_seq4	75.173	721	170	4	153	867	47	764	1.12E-135	484
	AR1_GAGATTCC-CCTATCCT_comp12530_seq0	78.772	1140	229	7	68	1202	54	1185	0	946
	YL3_CGCTCATT-ATAGAGGC_comp5037_seq1	71.826	827	212	10	321	1141	2325	3136	5.21E-115	416
	YL3_CGCTCATT-ATAGAGGC_comp5037_seq1	82	50	9	0	139	188	2143	2192	8.30E-05	50
	YS1_GAGATTCC-TATAGCCT_comp5438_seq0	71.814	816	210	9	332	1141	2238	3039	6.34E-114	412
	YS1_GAGATTCC-TATAGCCT_comp5438_seq0	82	50	9	0	139	188	2046	2095	8.30E-05	50
	AR1_GAGATTCC-CCTATCCT_comp2987_seq0	69.652	804	222	9	332	1127	359	1148	5.95E-89	329
D4H (U71604.1)	OL1_CGCTCATT-CCTATCCT_comp614_seq1	69.406	791	227	7	332	1119	5056	5834	1.08E-85	318
	AR1_GAGATTCC-CCTATCCT_comp10689_seq2	69.337	1298	322	24	31	1314	19	1254	1.18E-142	508
	AR2_GAGATTCC-GGCTCTGA_comp333_seq6	66.023	1295	364	23	31	1314	142	1371	4.12E-85	316
	AR2_GAGATTCC-GGCTCTGA_comp333_seq4	66.023	1295	364	23	31	1314	142	1371	4.12E-85	316
	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	66.486	555	168	7	202	753	575	1114	4.41E-34	147

TableS1

	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	69.082	207	52	3	1108	1314	1415	1609	1.54E-14	82.4
	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	82.857	35	6	0	41	75	54	88	0.57	37.4
DAT (AF053307.1)	AR2_GAGATTCC-GGCTCTGA_comp1084_seq0	63.105	1301	353	34	31	1314	402	1592	5.37E-33	143
	YS2_GAGATTCC-ATAGAGGC_comp528_seq2	69.342	1383	373	21	92	1438	2198	3565	9.78E-151	535
	YS1_GAGATTCC-TATAGCCT_comp15820_seq1	69.136	1377	388	15	94	1438	298	1669	3.41E-150	533
	YL1_CGCTCATT-TATAGCCT_comp104_seq6	69.27	1383	374	21	92	1438	2365	3732	4.16E-149	529
	OL3_CGCTCATT-GGCTCTGA_comp1489_seq0	66.935	1367	413	18	103	1446	919	2269	1.36E-104	381
TEX1 (MG873080.1)	AR2_GAGATTCC-GGCTCTGA_comp863_seq0	66.012	1324	427	12	124	1431	314	1630	9.18E-88	325
	YS2_GAGATTCC-ATAGAGGC_comp528_seq2	69.192	1373	394	15	103	1462	2209	3565	1.48E-148	527
	YL1_CGCTCATT-TATAGCCT_comp104_seq6	69.119	1373	395	15	103	1462	2376	3732	1.80E-147	524
	YS1_GAGATTCC-TATAGCCT_comp15820_seq1	67.632	1347	409	15	124	1455	328	1662	1.58E-116	421
	OL3_CGCTCATT-GGCTCTGA_comp1489_seq0	65.852	1391	412	17	103	1470	919	2269	2.86E-94	347
TEX2 (MG873081.1)	AR2_GAGATTCC-GGCTCTGA_comp863_seq0	65.799	1345	429	9	124	1460	314	1635	4.24E-92	340
	OL3_CGCTCATT-GGCTCTGA_comp1418_seq0	72.105	1520	382	16	1	1505	452	1944	0	796
	OL3_CGCTCATT-GGCTCTGA_comp12875_seq0	65.649	1278	401	21	97	1361	301	1553	7.56E-70	266
	AR1_GAGATTCC-CCTATCCT_comp10373_seq0	68.027	294	92	2	1054	1346	1064	1356	6.64E-20	100
	AR2_GAGATTCC-GGCTCTGA_comp863_seq0	73.81	168	39	3	1207	1370	1388	1554	8.09E-19	96.9
T19H (HQ901597.1)	OL1_CGCTCATT-CCTATCCT_comp520_seq1	70.918	196	56	1	1162	1357	1433	1627	9.86E-18	93.3
	AR1_GAGATTCC-CCTATCCT_comp10689_seq2	69.219	1293	330	19	1	1259	4	1262	2.75E-144	513
	AR2_GAGATTCC-GGCTCTGA_comp333_seq6	69.438	1263	329	22	14	1249	137	1369	3.13E-137	489
	AR2_GAGATTCC-GGCTCTGA_comp333_seq4	69.438	1263	329	22	14	1249	137	1369	3.13E-137	489
	AR2_GAGATTCC-GGCTCTGA_comp1084_seq0	64.969	1276	351	31	13	1261	396	1602	3.83E-60	233
	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	65.961	1087	317	26	190	1263	575	1621	6.93E-57	223
TAT (KU821123.1)	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	96.154	26	1	0	418	443	464	489	0.013	42.8
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	74.058	902	186	12	19	916	3631	2774	7.04E-159	562
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	81.435	237	44	0	916	1152	2478	2242	5.43E-59	230
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	88.889	108	12	0	1327	1434	2235	2128	2.17E-32	141
	YL3_CGCTCATT-ATAGAGGC_comp5333_seq0	77.317	723	152	4	195	916	3124	2413	8.57E-158	558
	YL3_CGCTCATT-ATAGAGGC_comp5333_seq0	79.032	372	65	3	1097	1461	2046	1681	1.36E-85	318
	YL3_CGCTCATT-ATAGAGGC_comp5333_seq0	81.356	59	11	0	916	974	2117	2059	7.09E-07	57.2
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	75.657	723	149	5	195	916	2711	2015	3.20E-144	513
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	85.149	202	23	1	1327	1521	1586	1385	1.56E-59	232
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	81.356	59	11	0	916	974	1719	1661	7.09E-07	57.2
	YS2_GAGATTCC-ATAGAGGC_comp8318_seq2	78.571	56	12	0	1097	1152	1648	1593	0.000367	48.2
	AR1_GAGATTCC-CCTATCCT_comp5822_seq1	71.865	622	161	7	868	1485	978	1589	1.66E-84	315
V19H (MK050464.10	YS1_GAGATTCC-TATAGCCT_comp6764_seq2	69.771	698	183	10	805	1488	2776	3459	2.31E-76	288
	AR1_GAGATTCC-CCTATCCT_comp10689_seq2	68.261	1317	321	24	27	1326	18	1254	2.98E-131	470
	AR2_GAGATTCC-GGCTCTGA_comp333_seq6	66.412	1307	354	18	28	1326	142	1371	1.77E-102	374
	AR2_GAGATTCC-GGCTCTGA_comp333_seq4	66.412	1307	354	18	28	1326	142	1371	1.77E-102	374
	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	68.467	574	163	7	190	760	566	1124	1.89E-51	205
	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	72.596	208	43	6	1120	1326	1415	1609	4.75E-21	104

	YS2_GAGATTCC-ATAGAGGC_comp13380_seq0	82.857	35	6	0	38	72	54	88	0.58	37.4
	AR2_GAGATTCC-GGCTCTGA_comp1084_seq0	66.725	568	147	12	28	592	402	930	8.59E-37	156
MAT (AF253415.1)	AR2_GAGATTCC-GGCTCTGA_comp1084_seq0	68.722	227	58	4	1102	1326	1377	1592	4.45E-15	84.2
	YL1_CGCTCATT-TATAGCCT_comp302_seq2	98.046	921	18	0	49	969	157	1077	0	1581
	YL1_CGCTCATT-TATAGCCT_comp302_seq2	83.488	969	160	0	1	969	1590	2558	0	1027
	AR2_GAGATTCC-GGCTCTGA_comp630_seq4	97.001	867	26	0	103	969	1291	2157	0	1447
	YS1_GAGATTCC-TATAGCCT_comp198_seq4	80.997	863	161	1	107	969	2373	3232	0	816
	AR1_GAGATTCC-CCTATCCT_comp3145_seq1	84.248	711	97	3	159	866	1	699	0	778
	AR1_GAGATTCC-CCTATCCT_comp3145_seq1	79.048	105	22	0	865	969	815	919	2.17E-17	91.5
VmPiNMT (KC70845(	OL3_CGCTCATT-GGCTCTGA_comp6929_seq0	77.855	867	192	0	103	969	218	1084	0	699
Transporters (TPs):											
TP	**transcript**	**% ID**	**Alignment length**	**Mismatch**	**Gapopen**	**TP start**	**TP end**	**Transcript start**	**Transcript end**	**E-value**	**Bitscore**
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	77.333	1725	381	6	192	1914	713	2429	0	1330
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	87.179	156	20	0	239	394	401	556	1.73E-47	192
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	78.431	51	11	0	192	242	160	210	0.02	42.8
	OL1_CGCTCATT-CCTATCCT_comp26741_seq0	76.236	1456	337	3	459	1914	1472	2918	0	1061
	YL1_CGCTCATT-TATAGCCT_comp4178_seq2	75.326	1459	351	3	459	1914	1452	2904	0	1003
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	77.114	887	200	1	1031	1914	2041	2927	0	682
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	76.014	838	195	2	202	1039	288	1119	1.01E-170	601
NPF2.1 (KR054375.1)	YS2_GAGATTCC-ATAGAGGC_comp11749_seq1	75.369	881	214	1	1034	1914	1179	2056	6.83E-173	609
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	77.493	1755	377	10	96	1844	683	2425	0	1355
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	83.117	154	26	0	173	326	401	554	2.82E-38	161
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	79.012	81	17	0	96	176	130	210	1.37E-10	69.8
	OL1_CGCTCATT-CCTATCCT_comp26741_seq0	76.823	1454	328	4	393	1846	1472	2916	0	1094
	YL1_CGCTCATT-TATAGCCT_comp4178_seq2	75.585	1454	352	2	393	1846	1452	2902	0	1016
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	75.312	960	214	7	18	973	179	1119	0	652
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	76.303	844	200	0	965	1808	2041	2884	1.04E-176	621
NPF2.2 (KR054376.1)	YS2_GAGATTCC-ATAGAGGC_comp11749_seq1	76.109	879	207	2	968	1846	1179	2054	5.77E-180	632
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	77.803	1784	366	6	128	1908	673	2429	0	1433
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	84.416	154	24	0	218	371	401	554	5.62E-41	170
	OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	72.165	97	18	3	128	221	120	210	0.000462	48.2
	YL1_CGCTCATT-TATAGCCT_comp4178_seq2	79.691	1423	277	2	438	1854	1452	2868	0	1265
	OL1_CGCTCATT-CCTATCCT_comp26741_seq0	79.534	1417	284	1	438	1854	1472	2882	0	1249
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	81.758	899	152	3	1010	1908	2041	2927	0	879
	YL1_CGCTCATT-TATAGCCT_comp3070_seq2	81.229	911	156	4	111	1018	221	1119	0	868
NPF2.3 (KR054377.1)	YS2_GAGATTCC-ATAGAGGC_comp11749_seq1	80.048	842	168	0	1013	1854	1179	2020	0	762
	OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	82.886	1788	291	6	1	1788	109	1881	0	1831
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	69.693	1630	418	20	106	1717	519	2090	0	691
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	78.218	303	66	0	79	381	185	487	6.84E-65	250
	YS2_GAGATTCC-ATAGAGGC_comp3442_seq2	68.611	1663	499	11	64	1716	335	1984	1.23E-175	618
	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	68.163	1671	492	11	55	1717	156	1794	9.46E-171	601

TableS1

	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	71.411	829	225	5	58	880	289	1111	7.78E-115	416
NPF2.4 (KR054378.1)	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	66.026	833	277	4	887	1716	1636	2465	4.92E-54	214
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	79.799	1688	286	9	86	1767	517	2155	0	1516
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	87.047	386	47	1	6	391	133	515	3.98E-131	470
	OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	70.662	1677	456	12	40	1701	163	1818	0	785
	YS2_GAGATTCC-ATAGAGGC_comp3442_seq2	69.952	1674	469	13	46	1705	335	1988	0	724
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	76.761	809	183	3	6	813	258	1062	2.68E-171	603
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	66.627	851	267	7	865	1705	1626	2469	1.22E-61	239
NPF2.5 (KR054379.1)	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	68.378	1714	487	22	40	1736	159	1834	1.14E-169	598
	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	86.18	1780	231	2	1	1780	93	1857	0	2105
	OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	68.645	1668	503	14	64	1718	160	1820	7.74E-172	605
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	68.741	1398	393	12	113	1499	517	1881	3.29E-151	536
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	74.32	331	83	2	56	385	153	482	7.28E-52	206
	YS2_GAGATTCC-ATAGAGGC_comp3442_seq2	67.629	1662	511	16	72	1717	334	1984	2.08E-147	524
NPF2.6 (KR054380.1)	YL1_CGCTCATT-TATAGCCT_comp5423_seq0	85.583	326	47	0	1	326	74	399	6.80E-103	376
	YS2_GAGATTCC-ATAGAGGC_comp3442_seq2	86.403	1743	234	1	37	1779	308	2047	0	2073
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	84.802	908	135	1	872	1779	1624	2528	0	1014
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	84.54	815	123	2	64	875	295	1109	0	895
	OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	69.424	1668	469	16	64	1712	169	1814	0	673
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	68.54	1548	425	12	104	1640	517	2013	1.15E-169	598
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	78.246	285	62	0	64	348	170	454	1.50E-60	235
NPF2.7 (KR054381.1)	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	67.351	1657	503	12	63	1712	164	1789	4.88E-149	529
	YS2_GAGATTCC-ATAGAGGC_comp3442_seq2	79.965	1727	330	6	76	1797	332	2047	0	1543
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	77.133	914	197	4	887	1797	1624	2528	0	699
	YL3_CGCTCATT-ATAGAGGC_comp3976_seq4	79.075	822	161	5	79	893	295	1112	0	695
	OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	66.707	1670	515	19	79	1729	169	1816	2.55E-127	457
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	66.338	1625	479	18	119	1729	517	2087	1.61E-123	444
	YL3_CGCTCATT-ATAGAGGC_comp13973_seq0	75.09	277	67	2	88	363	179	454	2.93E-44	181
	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	69.159	963	265	19	78	1027	164	1107	1.42E-92	342
NPF2.8 (KR054382.1)	AR2_GAGATTCC-GGCTCTGA_comp6824_seq0	63.089	382	126	3	1387	1768	1464	1830	1.02E-05	53.6
	YL3_CGCTCATT-ATAGAGGC_comp2760_seq1	78.218	1717	351	5	1	1716	254	1948	0	1409
	YL1_CGCTCATT-TATAGCCT_comp789_seq5	74.941	850	198	2	867	1716	3127	3961	1.03E-163	578
	YL1_CGCTCATT-TATAGCCT_comp789_seq6	74.706	850	200	2	867	1716	3966	4800	5.36E-161	569
	YL1_CGCTCATT-TATAGCCT_comp789_seq0	74.706	850	200	2	867	1716	3922	4756	5.36E-161	569
	YL1_CGCTCATT-TATAGCCT_comp28142_seq0	74.441	313	80	0	1404	1716	158	470	2.45E-51	205
	YL1_CGCTCATT-TATAGCCT_comp28142_seq0	83.803	142	23	0	829	970	15	156	1.35E-35	152
NPF2.9 (KX372303.1)	YL1_CGCTCATT-TATAGCCT_comp28142_seq0	76	75	18	0	1642	1716	495	569	2.80E-06	55.4
	AR1_GAGATTCC-CCTATCCT_comp18788_seq0	82.497	1554	260	3	1	1554	192	1733	0	1573
	OL1_CGCTCATT-CCTATCCT_comp74_seq2	81.778	1575	275	3	1	1575	2257	3819	0	1544
	YS1_GAGATTCC-TATAGCCT_comp122_seq0	81.651	1575	277	3	1	1575	1509	3071	0	1535
	OL3_CGCTCATT-GGCTCTGA_comp1979_seq2	81.587	1575	278	3	1	1575	936	2498	0	1530

MATE1 (KX372304.1)	OL1_CGCTCATT-CCTATCCT_comp7876_seq0	83.482	1235	204	0	320	1554	3127	4361	0	1308
	YS2_GAGATTCC-ATAGAGGC_comp3345_seq1	82.43	1383	237	3	112	1491	308	1687	0	1391
	YS2_GAGATTCC-ATAGAGGC_comp3345_seq1	85.714	35	5	0	9	43	229	263	0.053	41
	YS1_GAGATTCC-TATAGCCT_comp124_seq0	82.43	1383	237	3	112	1491	2469	3848	0	1391
	YS1_GAGATTCC-TATAGCCT_comp124_seq0	84.615	39	6	0	5	43	2386	2424	0.004	44.6
	YL3_CGCTCATT-ATAGAGGC_comp598_seq6	82.357	1383	238	3	112	1491	2895	4274	0	1386
	YL3_CGCTCATT-ATAGAGGC_comp598_seq6	84.615	39	6	0	5	43	2812	2850	0.004	44.6
	YS2_GAGATTCC-ATAGAGGC_comp3345_seq7	82.357	1383	238	3	112	1491	311	1690	0	1386
	YS2_GAGATTCC-ATAGAGGC_comp3345_seq7	85.714	35	5	0	9	43	229	263	0.053	41
	AR1_GAGATTCC-CCTATCCT_comp14782_seq0	76.637	672	154	1	820	1491	941	1609	5.66E-141	502
MATE2 (KX372305.1)	AR1_GAGATTCC-CCTATCCT_comp14782_seq0	80	40	8	0	320	359	441	480	0.65	37.4
	OL3_CGCTCATT-GGCTCTGA_comp2267_seq1	99.184	4044	32	1	1	4044	1547	5589	0	7142
	YL1_CGCTCATT-TATAGCCT_comp818_seq3	78.995	1771	372	0	2274	4044	2892	4662	0	1517
	YL1_CGCTCATT-TATAGCCT_comp818_seq3	70.379	2245	628	17	40	2270	593	2814	0	1009
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	71.981	2873	768	17	40	2898	579	3428	0	1510
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	75.32	859	212	0	2908	3766	3815	4673	3.20E-168	594
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	82.927	82	14	0	3771	3852	4935	5016	3.93E-15	86
	YS1_GAGATTCC-TATAGCCT_comp1672_seq3	71.911	2873	770	17	40	2898	608	3457	0	1501
	YS1_GAGATTCC-TATAGCCT_comp1672_seq3	75.661	945	230	0	2908	3852	3901	4845	0	668
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	71.536	2252	613	12	655	2898	181	2412	0	1144
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	75.443	790	194	0	2977	3766	3027	3816	3.42E-155	551
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	81.707	82	15	0	3771	3852	4078	4159	1.67E-13	80.6
VmTPT2 (KC511773.1	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	80.263	76	15	0	2908	2983	2853	2928	3.02E-10	69.8
	OL3_CGCTCATT-GGCTCTGA_comp2267_seq1	87.305	4230	534	1	37	4266	1363	5589	0	5205
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	71.399	2867	780	17	268	3120	588	3428	0	1433
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	75.086	871	215	2	3119	3988	3804	4673	1.75E-165	585
	YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	81.053	95	18	0	3993	4087	4935	5029	9.75E-17	91.5
	YS1_GAGATTCC-TATAGCCT_comp1672_seq3	71.259	2867	784	17	268	3120	617	3457	0	1415
	YS1_GAGATTCC-TATAGCCT_comp1672_seq3	75.67	970	234	2	3119	4087	3890	4858	0	679
	YL1_CGCTCATT-TATAGCCT_comp818_seq3	77.314	1772	400	2	2496	4266	2892	4662	0	1377
	YL1_CGCTCATT-TATAGCCT_comp818_seq3	71.251	2014	562	9	268	2270	602	2609	0	996
	YL1_CGCTCATT-TATAGCCT_comp818_seq3	75	68	17	0	2425	2492	2747	2814	0.004	46.4
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	71.188	2256	617	15	874	3120	181	2412	0	1101
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	75.601	791	191	2	3199	3988	3027	3816	1.26E-154	549
	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	88.06	67	8	0	3993	4059	4078	4144	4.14E-15	86
CrTPT2 (KC511771.1)	OL1_CGCTCATT-CCTATCCT_comp1940_seq1	77.011	87	20	0	3119	3205	2842	2928	1.11E-09	68
Transcription Factors (TFs):											
TF	**transcript**	**% ID**	**Alignment length**	**Mismatch**	**Gapopen**	**TF start**	**TF end**	**Transcript start**	**Transcript end**	**E-value**	**Bitscore**
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	78.509	228	49	0	365	592	662	889	1.93E-47	190
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	74.308	253	59	2	365	611	688	940	3.27E-38	159

TableS1

	YS2_GAGATTCC-ATAGAGGC_comp27339_seq0	74.336	226	45	1	367	592	422	634	1.39E-36	154
	YS2_GAGATTCC-ATAGAGGC_comp27339_seq0	91.429	35	3	0	76	110	134	168	4.27E-05	50
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	75.568	176	43	0	376	551	496	671	2.36E-27	123
ORCA2 (AJ238740.1)	YS1_GAGATTCC-TATAGCCT_comp5858_seq0	73.988	173	45	0	379	551	616	788	5.19E-23	109
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	80.669	269	49	1	1085	1353	666	931	6.95E-65	250
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	76.316	76	18	0	837	912	124	199	8.49E-07	57.2
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	84.615	182	28	0	1122	1303	497	678	9.05E-51	203
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	78.947	209	44	0	1104	1312	656	864	1.03E-43	179
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	78.481	79	14	2	823	898	33	111	6.97E-08	60.8
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	75	48	12	0	991	1038	345	392	9.7	33.7
	AR1_GAGATTCC-CCTATCCT_comp14477_seq0	83.333	138	23	0	1112	1249	458	595	2.13E-33	145
ORCA3 (AJ251250.1)	AR1_GAGATTCC-CCTATCCT_comp11153_seq4	73.14	242	62	1	1145	1386	435	673	2.59E-32	141
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	78.516	256	55	0	340	595	675	930	1.80E-54	214
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	81.928	83	10	3	16	95	126	206	3.98E-12	73.4
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	84.066	182	29	0	365	546	497	678	1.39E-49	197
	AR1_GAGATTCC-CCTATCCT_comp11153_seq4	77.542	236	51	2	361	595	408	642	3.72E-44	179
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	74.429	219	55	1	337	554	645	863	3.26E-32	140
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	86.667	60	8	0	19	78	55	114	3.98E-12	73.4
ORCA4 (KR703577.1	AR1_GAGATTCC-CCTATCCT_comp14477_seq0	80.142	141	28	0	355	495	458	598	5.90E-29	129
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	81.191	319	51	2	614	932	488	797	1.83E-81	304
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	78.75	80	14	1	25	101	178	257	8.33E-10	66.2
	AR1_GAGATTCC-CCTATCCT_comp14477_seq0	81.503	173	29	1	584	753	426	598	3.31E-40	167
	AR1_GAGATTCC-CCTATCCT_comp14477_seq0	80.723	83	10	1	25	101	146	228	1.32E-13	78.8
	YL1_CGCTCATT-TATAGCCT_comp6621_seq0	81.437	167	31	0	631	797	708	874	1.41E-38	161
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	81.679	131	24	0	631	761	682	812	8.32E-29	129
ORCA5 (KR703578.1)	YL1_CGCTCATT-TATAGCCT_comp10613_seq0	72.989	174	47	0	624	797	554	727	1.16E-20	102
	AR2_GAGATTCC-GGCTCTGA_comp3513_seq0	82.162	185	33	0	271	455	487	671	7.12E-46	185
	AR1_GAGATTCC-CCTATCCT_comp14477_seq0	80.952	147	28	0	271	417	458	604	2.66E-32	140
	OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	74.869	191	48	0	281	471	674	864	4.80E-29	129
	YS1_GAGATTCC-TATAGCCT_comp5858_seq0	74.566	173	44	0	283	455	616	788	1.06E-24	114
	YS2_GAGATTCC-ATAGAGGC_comp27339_seq0	70.244	205	48	2	267	471	418	609	8.13E-20	98.7
ORCA6 (MN614455.1	YS2_GAGATTCC-ATAGAGGC_comp27339_seq0	85.714	28	4	0	126	153	235	262	2.8	33.7
	YL3_CGCTCATT-ATAGAGGC_comp4742_seq3	73.982	1130	199	22	3	1101	269	1334	0	699
	AR1_GAGATTCC-CCTATCCT_comp8525_seq2	74.291	564	133	4	511	1071	750	1304	3.90E-97	356
	AR1_GAGATTCC-CCTATCCT_comp8525_seq2	78.947	114	19	2	7	116	285	397	7.07E-18	93.3
	OL1_CGCTCATT-CCTATCCT_comp6055_seq0	65.82	433	128	7	622	1038	880	1308	4.76E-20	100
	AR2_GAGATTCC-GGCTCTGA_comp1597_seq3	78.947	114	19	2	7	116	300	412	7.07E-18	93.3
BIS1 (KM409646.2)	YL1_CGCTCATT-TATAGCCT_comp10816_seq0	93.548	31	1	1	859	888	1680	1650	0.003	44.6
	AR1_GAGATTCC-CCTATCCT_comp8525_seq2	70.934	867	152	15	372	1230	551	1325	2.60E-125	450
	AR1_GAGATTCC-CCTATCCT_comp8525_seq2	69.811	159	24	4	1	159	282	416	2.15E-12	75.2
	YL3_CGCTCATT-ATAGAGGC_comp4742_seq3	71.575	584	142	5	631	1214	750	1309	2.44E-81	304

TableS1

	YL3_CGCTCATT-ATAGAGGC_comp4742_seq3	72.857	70	7	3	9	78	275	332	0.15	39.2
	OL1_CGCTCATT-CCTATCCT_comp6055_seq0	71.171	222	64	0	662	883	809	1030	8.54E-24	113
	OL1_CGCTCATT-CCTATCCT_comp6055_seq0	67.485	163	53	0	1013	1175	1145	1307	2.01E-06	55.4
	AR2_GAGATTCC-GGCTCTGA_comp1597_seq3	69.811	159	24	4	1	159	297	431	2.15E-12	75.2
	AR2_GAGATTCC-GGCTCTGA_comp1597_seq3	76	100	15	3	372	468	566	659	3.19E-10	68
	AR2_GAGATTCC-GGCTCTGA_comp1597_seq3	83.871	31	5	0	215	245	466	496	6.6	33.7
BIS2 (KM409645.1)	AR2_GAGATTCC-GGCTCTGA_comp1660_seq1	80.952	63	12	0	757	819	891	953	4.73E-08	60.8
	OL3_CGCTCATT-GGCTCTGA_comp2316_seq0	80	1415	181	27	1158	2552	1441	2773	0	1247
	OL1_CGCTCATT-CCTATCCT_comp4812_seq0	80.783	1353	191	24	1185	2501	1078	2397	0	1222
	OL1_CGCTCATT-CCTATCCT_comp4812_seq0	70.926	1080	191	21	77	1076	8	1044	7.71E-155	549
	OL1_CGCTCATT-CCTATCCT_comp2168_seq4	74.927	1029	181	18	77	1079	107	1084	0	681
	YS1_GAGATTCC-TATAGCCT_comp2461_seq6	74.636	1029	184	18	77	1079	204	1181	0	668
	YS1_GAGATTCC-TATAGCCT_comp14327_seq2	80.18	666	91	18	1895	2540	3	647	1.22E-158	562
	YS1_GAGATTCC-TATAGCCT_comp14327_seq2	76.087	92	8	5	2261	2350	655	734	1.23E-06	57.2
MYC2 (AF283507.2)	YS1_GAGATTCC-TATAGCCT_comp14327_seq2	73.529	102	16	5	2450	2547	735	829	0.000183	50
	AR1_GAGATTCC-CCTATCCT_comp24292_seq0	77.204	329	51	6	322	632	542	864	2.93E-65	250
	YS1_GAGATTCC-TATAGCCT_comp87_seq4	71.498	414	85	8	30	425	154	552	8.96E-53	208
	YS1_GAGATTCC-TATAGCCT_comp87_seq3	71.294	425	83	10	30	427	154	566	3.13E-52	206
	AR1_GAGATTCC-CCTATCCT_comp16215_seq2	77.533	227	30	7	424	632	808	1031	2.74E-40	167
WRKY1 (HQ646368.1	AR2_GAGATTCC-GGCTCTGA_comp1658_seq0	75.771	227	31	6	424	632	3831	4051	4.96E-37	156
	AR1_GAGATTCC-CCTATCCT_comp737_seq2	72.233	533	94	14	104	600	328	842	1.76E-74	280
	AR1_GAGATTCC-CCTATCCT_comp737_seq2	96.97	33	1	0	70	102	267	299	1.27E-06	55.4
	YL1_CGCTCATT-TATAGCCT_comp10081_seq1	69.458	203	56	2	206	405	179	378	2.62E-15	84.2
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq0	70.408	196	53	4	203	397	261	452	1.11E-13	78.8
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq2	68.837	215	52	5	194	396	240	451	3.89E-13	77
ZCT1 (AJ632082.1)	AR1_GAGATTCC-CCTATCCT_comp6246_seq5	78.75	80	15	2	200	278	527	605	2.99E-08	60.8
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq0	69.737	608	124	15	80	657	90	667	3.09E-66	253
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq0	73.75	160	29	9	805	958	840	992	2.09E-11	71.6
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq2	66.026	624	111	19	64	657	76	628	6.82E-43	176
	AR2_GAGATTCC-GGCTCTGA_comp7502_seq2	71.519	158	27	9	808	958	802	948	1.32E-07	59
	AR1_GAGATTCC-CCTATCCT_comp737_seq2	71.689	219	51	5	236	447	409	623	3.54E-21	104
	AR2_GAGATTCC-GGCTCTGA_comp5665_seq0	80.556	72	14	0	379	450	1288	1359	2.55E-10	68
	AR2_GAGATTCC-GGCTCTGA_comp5665_seq0	71.053	76	22	0	238	313	1096	1171	0.12	39.2
ZCT2 (AJ632083.1)	YL1_CGCTCATT-TATAGCCT_comp10081_seq1	73.737	99	26	0	214	312	145	243	1.08E-08	62.6
	AR1_GAGATTCC-CCTATCCT_comp521_seq0	72.048	991	174	23	61	1023	171	1086	9.68E-149	527
	AR2_GAGATTCC-GGCTCTGA_comp1098_seq4	71.287	1010	166	28	61	1019	1556	2492	1.35E-140	500
	OL3_CGCTCATT-GGCTCTGA_comp7752_seq4	71.125	987	179	25	57	998	186	1111	4.40E-134	479
	OL1_CGCTCATT-CCTATCCT_comp2136_seq4	69.679	343	63	13	676	1005	926	1240	9.12E-29	129
ZCT3 (AJ632084.1)	AR2_GAGATTCC-GGCTCTGA_comp10369_seq0	75.862	87	15	1	679	765	481	401	3.19E-09	64.4
	YL1_CGCTCATT-TATAGCCT_comp1354_seq1	68.293	1066	143	29	1	1042	103	997	9.27E-124	444
	YS2_GAGATTCC-ATAGAGGC_comp2064_seq8	67.864	1058	146	26	1	1042	2965	3844	1.68E-120	434

206

	OL3_CGCTCATT-GGCTCTGA_comp8246_seq1	67.628	1075	149	31	1	1044	49	955	1.91E-113	410
	YS1_GAGATTCC-TATAGCCT_comp2284_seq2	68.045	1064	149	35	1	1042	119	1013	9.91E-111	401
	YS2_GAGATTCC-ATAGAGGC_comp4735_seq0	73.786	515	81	9	539	1044	326	795	5.14E-89	329
GATA1 (MK801106.1	YS2_GAGATTCC-ATAGAGGC_comp4735_seq0	91.304	23	2	0	83	105	181	203	5.5	33.7
	YS1_GAGATTCC-TATAGCCT_comp3725_seq0	77.524	1882	300	38	5	1858	186	1972	0	1427
	YS1_GAGATTCC-TATAGCCT_comp3725_seq0	100	19	0	0	1	19	168	186	2.9	35.6
	YL1_CGCTCATT-TATAGCCT_comp4560_seq1	80.074	1345	215	17	5	1320	130	1450	0	1189
	YL1_CGCTCATT-TATAGCCT_comp4560_seq1	72.176	478	66	20	1386	1858	1716	2131	5.51E-60	233
	YL1_CGCTCATT-TATAGCCT_comp4560_seq1	100	19	0	0	1	19	112	130	2.9	35.6
	YL1_CGCTCATT-TATAGCCT_comp4560_seq0	78.753	1219	199	12	6	1205	112	1289	0	1031
	YL1_CGCTCATT-TATAGCCT_comp4560_seq0	71.084	664	107	22	1203	1858	1393	1979	8.73E-83	309
	OL1_CGCTCATT-CCTATCCT_comp4376_seq4	81.124	694	116	2	621	1314	849	1527	0	666
	OL1_CGCTCATT-CCTATCCT_comp4376_seq4	77.671	627	98	10	6	613	125	728	3.70E-138	493
	OL1_CGCTCATT-CCTATCCT_comp4376_seq4	69.933	449	78	17	1380	1825	1668	2062	1.30E-42	176
GBF1 (AF084971.1)	YL1_CGCTCATT-TATAGCCT_comp4162_seq2	75.269	93	23	0	1118	1210	1082	1174	5.90E-09	64.4
	OL3_CGCTCATT-GGCTCTGA_comp5631_seq2	81.358	1797	232	28	22	1786	124	1849	0	1698
	AR1_GAGATTCC-CCTATCCT_comp6026_seq0	83.374	1215	146	18	574	1769	706	1883	0	1251
	AR1_GAGATTCC-CCTATCCT_comp6026_seq0	85.985	264	32	2	355	618	398	656	1.08E-81	306
	AR1_GAGATTCC-CCTATCCT_comp6026_seq0	69.424	278	48	10	22	289	138	388	5.61E-22	107
	AR2_GAGATTCC-GGCTCTGA_comp6592_seq4	82.512	1258	152	21	574	1803	761	1978	0	1243
	AR2_GAGATTCC-GGCTCTGA_comp6592_seq4	77.758	571	78	14	22	582	146	677	7.27E-122	439
	YS1_GAGATTCC-TATAGCCT_comp5460_seq5	84.724	995	119	8	188	1163	234	1214	0	1103
	YS1_GAGATTCC-TATAGCCT_comp5460_seq5	80.787	635	79	15	1160	1786	2377	2976	2.08E-160	567
	YS1_GAGATTCC-TATAGCCT_comp5460_seq5	73.196	97	23	2	5	100	83	177	3.79E-05	51.8
	YS1_GAGATTCC-TATAGCCT_comp5460_seq5	81.818	44	8	0	82	125	180	223	0.006	44.6
	AR1_GAGATTCC-CCTATCCT_comp11579_seq0	82.381	840	109	15	866	1686	1895	2714	0	816
	AR1_GAGATTCC-CCTATCCT_comp11579_seq0	92.069	290	23	0	574	863	738	1027	6.81E-116	419
GBF2 (AF084972.1)	AR1_GAGATTCC-CCTATCCT_comp11579_seq0	75.207	605	81	16	18	582	78	653	1.50E-111	405
	AR2_GAGATTCC-GGCTCTGA_comp630_seq3	78.341	651	78	11	385	1026	424	1020	3.62E-154	545
	AR2_GAGATTCC-GGCTCTGA_comp630_seq3	82.692	104	15	1	1	101	157	260	1.04E-21	105
	AR2_GAGATTCC-GGCTCTGA_comp630_seq3	82.051	78	14	0	236	313	323	400	1.45E-13	78.8
	AR2_GAGATTCC-GGCTCTGA_comp630_seq4	78.341	651	78	11	385	1026	424	1020	3.62E-154	545
	AR2_GAGATTCC-GGCTCTGA_comp630_seq4	82.692	104	15	1	1	101	157	260	1.04E-21	105
	AR2_GAGATTCC-GGCTCTGA_comp630_seq4	82.051	78	14	0	236	313	323	400	1.45E-13	78.8
	AR2_GAGATTCC-GGCTCTGA_comp630_seq1	78.221	652	79	11	384	1026	420	1017	1.26E-153	544
	AR2_GAGATTCC-GGCTCTGA_comp630_seq1	82.692	104	15	1	1	101	157	260	1.04E-21	105
	AR2_GAGATTCC-GGCTCTGA_comp630_seq1	80	95	19	0	236	330	323	417	9.77E-16	86
	OL1_CGCTCATT-CCTATCCT_comp1173_seq1	78.034	651	80	11	385	1026	483	1079	1.88E-151	536
	OL1_CGCTCATT-CCTATCCT_comp1173_seq1	82.692	104	15	1	1	101	219	322	1.04E-21	105
	OL1_CGCTCATT-CCTATCCT_comp1173_seq1	79.487	78	16	0	236	313	382	459	7.51E-11	69.8
	YL3_CGCTCATT-ATAGAGGC_comp915_seq2	77.914	652	81	11	384	1026	630	1227	6.55E-151	535

CrERF5 (MK862158.1	YL3_CGCTCATT-ATAGAGGC_comp915_seq2	80.769	78	15	0	253	330	550	627	6.17E-12	73.4

TableS2

Transcript ID	MIA	Communities	sprot_Top_BLASTX_hit
YS2_GAGATTCC-ATAGAGGC_comp13351_seq1	NA	1	Y3037_ARATH^Y3037_ARATH^Q:412-2247,H:43-639^42.86%ID^E:4e-128^RecName: Full=Probable inactive leucine-rich repeat receptor-like protein kinase At3g03770
OL3_CGCTCATT-GGCTCTGA_comp10301_seq0	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp2586_seq1	NA	1	.
YL3_CGCTCATT-ATAGAGGC_comp19793_seq1	NA	1	C86A2_ARATH^C86A2_ARATH^Q:2-346,H:409-523^75.65%ID^E:2e-53^RecName: Full=Cytochrome P450 86A2
YS2_GAGATTCC-ATAGAGGC_comp1135_seq0	NA	1	RL212_ARATH^RL212_ARATH^Q:27-518,H:1-164^84.76%ID^E:7e-100^RecName: Full=60S ribosomal protein L21-2
AR1_GAGATTCC-CCTATCCT_comp8017_seq0	NA	1	DST1_DICDI^DST1_DICDI^Q:863-1696,H:15-283^57.55%ID^E:2e-97^RecName: Full=Serine/threonine-protein kinase dst1 {ECO:0000250|UniProtKB:O61125}
YL3_CGCTCATT-ATAGAGGC_comp3883_seq1	NA	1	F135B_XENLA^F135B_XENLA^Q:2876-3694,H:1102-1362^41.09%ID^E:5e-51^RecName: Full=Protein FAM135B
OL1_CGCTCATT-CCTATCCT_comp1577_seq4	NA	1	GLGL1_SOLTU^GLGL1_SOLTU^Q:941-2191,H:2-417^81.29%ID^E:0^RecName: Full=Glucose-1-phosphate adenylyltransferase large subunit 1
YS2_GAGATTCC-ATAGAGGC_comp561_seq1	NA	1	C81F1_ARATH^C81F1_ARATH^Q:172-1560,H:39-499^40.88%ID^E:3e-118^RecName: Full=Cytochrome P450 81F1 {ECO:0000305}
YS1_GAGATTCC-TATAGCCT_comp8849_seq1	NA	1	FDL1_ARATH^FDL1_ARATH^Q:168-1520,H:3-415^28.14%ID^E:1e-29^RecName: Full=F-box/FBD/LRR-repeat protein At1g13570
YL3_CGCTCATT-ATAGAGGC_comp188_seq0	NA	1	BGL17_ARATH^BGL17_ARATH^Q:1932-3023,H:33-395^56.87%ID^E:4e-135^RecName: Full=Beta-glucosidase 17
YS1_GAGATTCC-TATAGCCT_comp2178_seq1	NA	1	.
OL1_CGCTCATT-CCTATCCT_comp2611_seq3	NA	1	.
YL3_CGCTCATT-ATAGAGGC_comp7990_seq0	NA	1	EDL17_ARATH^EDL17_ARATH^Q:114-1505,H:10-467^48.08%ID^E:1e-128^RecName: Full=Sugar transporter ERD6-like 17
YL1_CGCTCATT-TATAGCCT_comp6895_seq1	NA	1	EAAC_ARATH^EAAC_ARATH^Q:83-1255,H:1-379^69.57%ID^E:1e-172^RecName: Full=Probable envelope ADP,ATP carrier protein, chloroplastic
AR2_GAGATTCC-GGCTCTGA_comp2954_seq0	NA	1	C71DC_CATRO^C71DC_CATRO^Q:267-1745,H:2-492^50.2%ID^E:1e-166^RecName: Full=Tabersonine 16-hydroxylase
YL1_CGCTCATT-TATAGCCT_comp3031_seq4	NA	1	.
YL3_CGCTCATT-ATAGAGGC_comp22491_seq0	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp15661_seq0	NA	1	.
YS1_GAGATTCC-TATAGCCT_comp3802_seq2	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp5672_seq0	NA	1	BH048_ARATH^BH048_ARATH^Q:632-1372,H:27-270^50.2%ID^E:3e-51^RecName: Full=Transcription factor bHLH48
YL1_CGCTCATT-TATAGCCT_comp3021_seq0	NA	1	.
YS2_GAGATTCC-ATAGAGGC_comp776_seq0	PAS	1	THCAS_CANSA^THCAS_CANSA^Q:197-1681,H:32-540^44.92%ID^E:1e-133^RecName: Full=Tetrahydrocannabinolic acid synthase
YS2_GAGATTCC-ATAGAGGC_comp1760_seq0	CrTPT2	1	AB32G_ARATH^AB32G_ARATH^Q:318-3440,H:1-1043^75.84%ID^E:0^RecName: Full=ABC transporter G family member 32 {ECO:0000303|PubMed:18299247}
YL3_CGCTCATT-ATAGAGGC_comp16614_seq0	NA	1	.
OL1_CGCTCATT-CCTATCCT_comp3680_seq3	NA	1	GLGL2_SOLTU^GLGL2_SOLTU^Q:3403-5007,H:1-519^73.83%ID^E:0^RecName: Full=Glucose-1-phosphate adenylyltransferase large subunit 2, chloroplastic/amyloplastic
YL1_CGCTCATT-TATAGCCT_comp12228_seq3	NA	1	.
YS2_GAGATTCC-ATAGAGGC_comp16202_seq0	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp2644_seq3	NA	1	GPDH2_ORYSJ^GPDH2_ORYSJ^Q:731-2029,H:16-382^68.13%ID^E:0^RecName: Full=Probable glycerol-3-phosphate dehydrogenase [NAD(+)] 2, cytosolic
YS2_GAGATTCC-ATAGAGGC_comp8501_seq0	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp33140_seq0	NA	1	AAE13_ARATH^AAE13_ARATH^Q:2119-3087,H:222-540^73.37%ID^E:2e-159^RecName: Full=Malonate--CoA ligase
OL1_CGCTCATT-CCTATCCT_comp517_seq13	NA	1	ECR_ARATH^ECR_ARATH^Q:4949-5878,H:1-310^81.29%ID^E:7e-174^.^.
AR2_GAGATTCC-GGCTCTGA_comp2244_seq1	STR	1	STSY_RAUSE^STSY_RAUSE^Q:58-1059,H:1-333^77.61%ID^E:3e-171^RecName: Full=Strictosidine synthase
YL1_CGCTCATT-TATAGCCT_comp2609_seq5	NA	1	.
OL1_CGCTCATT-CCTATCCT_comp3318_seq1	NA	1	CER1_ARATH^CER1_ARATH^Q:892-2229,H:171-620^59.87%ID^E:3e-159^RecName: Full=Protein ECERIFERUM 1
YL3_CGCTCATT-ATAGAGGC_comp6907_seq0	NA	1	Y1491_ARATH^Y1491_ARATH^Q:219-1655,H:21-499^78.29%ID^E:0^RecName: Full=Uncharacterized protein At1g04910
YS1_GAGATTCC-TATAGCCT_comp65_seq1	NA	1	MER_ERYCB^MER_ERYCB^Q:109-663,H:4-194^56.25%ID^E:7e-61^RecName: Full=Methylecgonone reductase
YL3_CGCTCATT-ATAGAGGC_comp3908_seq3	NA	1	CSE_ARATH^CSE_ARATH^Q:258-1172,H:23-323^31.7%ID^E:1e-41^RecName: Full=Caffeoylshikimate esterase
YL3_CGCTCATT-ATAGAGGC_comp1258_seq0	SAT	1	VINSY_RAUSE^VINSY_RAUSE^Q:163-1392,H:1-408^65.06%ID^E:0^RecName: Full=Vinorine synthase
YS2_GAGATTCC-ATAGAGGC_comp32416_seq0	NA	1	CADH6_ORYSJ^CADH6_ORYSJ^Q:13-246,H:70-147^75.64%ID^E:5e-34^RecName: Full=Probable cinnamyl alcohol dehydrogenase 6
YL3_CGCTCATT-ATAGAGGC_comp13569_seq0	NA	1	.
YS2_GAGATTCC-ATAGAGGC_comp3900_seq3	NA	1	FH3_ORYSJ^FH3_ORYSJ^Q:2023-3180,H:809-1205^77.58%ID^E:0^RecName: Full=Formin-like protein 3
YL1_CGCTCATT-TATAGCCT_comp193_seq6	NA	1	RL362_ARATH^RL362_ARATH^Q:123-452,H:1-112^80.36%ID^E:9e-49^RecName: Full=60S ribosomal protein L36-2

YS2_GAGATTCC-ATAGAGGC_comp2915_seq1	CMS	1	ISPD_ARATH^ISPD_ARATH^Q:1565-2311,H:56-299^77.51%ID^E:1e-121^RecName: Full=2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp5426_seq0	NA	1	TRY_ARATH^TRY_ARATH^Q:78-326,H:7-87^68.67%ID^E:1e-28^RecName: Full=Transcription factor TRY
YL1_CGCTCATT-TATAGCCT_comp6500_seq0	NA	1	SHN2_ARATH^SHN2_ARATH^Q:172-756,H:1-189^57.36%ID^E:2e-48^RecName: Full=Ethylene-responsive transcription factor SHINE 2
YL1_CGCTCATT-TATAGCCT_comp621_seq0	NA	1	.
YS1_GAGATTCC-TATAGCCT_comp445_seq2	DXR	1	DXR_MENPI^DXR_MENPI^Q:367-1776,H:1-470^86.86%ID^E:0^RecName: Full=1-deoxy-D-xylulose 5-phosphate reductoisomerase, chloroplastic
YS1_GAGATTCC-TATAGCCT_comp24972_seq0	SLS1	1	C72A1_CATRO^C72A1_CATRO^Q:1427-2701,H:93-517^91.29%ID^E:0^RecName: Full=Secologanin synthase
YL1_CGCTCATT-TATAGCCT_comp14134_seq1	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp15133_seq0	NA	1	GDL17_ARATH^GDL17_ARATH^Q:182-1192,H:34-370^70.21%ID^E:4e-171^RecName: Full=GDSL esterase/lipase At1g33811
YL1_CGCTCATT-TATAGCCT_comp2275_seq4	NA	1	UBP1_NICPL^UBP1_NICPL^Q:1100-2194,H:21-384^86.89%ID^E:0^RecName: Full=Oligouridylate-binding protein 1
YS2_GAGATTCC-ATAGAGGC_comp2920_seq0	NA	1	PPR39_ARATH^PPR39_ARATH^Q:245-658,H:69-204^39.13%ID^E:1e-35^RecName: Full=Pentatricopeptide repeat-containing protein At1g12775, mitochondrial
YS2_GAGATTCC-ATAGAGGC_comp10625_seq1	NA	1	Y2921_ARATH^Y2921_ARATH^Q:31-1602,H:4-502^32%ID^E:9e-60^RecName: Full=Putative leucine-rich repeat receptor-like protein kinase At2g19210
YL1_CGCTCATT-TATAGCCT_comp29020_seq2	NA	1	.
YS1_GAGATTCC-TATAGCCT_comp22171_seq0	NA	1	C81F1_ARATH^C81F1_ARATH^Q:852-1457,H:300-499^50%ID^E:1e-43^RecName: Full=Cytochrome P450 81F1 {ECO:0000305}
YL1_CGCTCATT-TATAGCCT_comp4005_seq1	NA	1	WSD1_ARATH^WSD1_ARATH^Q:319-1278,H:61-386^27.03%ID^E:1e-26^RecName: Full=O-acyltransferase WSD1
YS2_GAGATTCC-ATAGAGGC_comp3651_seq3	NA	1	FHYRK_ARATH^FHYRK_ARATH^Q:1731-2336,H:178-379^81.19%ID^E:6e-110^RecName: Full=Bifunctional riboflavin kinase/FMN phosphatase
YS2_GAGATTCC-ATAGAGGC_comp3692_seq0	NA	1	SCRK6_ARATH^SCRK6_ARATH^Q:420-1502,H:28-384^81.16%ID^E:1e-177^RecName: Full=Probable fructokinase-6, chloroplastic
YS1_GAGATTCC-TATAGCCT_comp18331_seq0	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp1075_seq1	NA	1	.
YL1_CGCTCATT-TATAGCCT_comp444_seq0	DL7H	1	C72A1_CATRO^C72A1_CATRO^Q:232-1896,H:1-522^51.17%ID^E:0^RecName: Full=Secologanin synthase
YL3_CGCTCATT-ATAGAGGC_comp9604_seq2	NA	1	KGLT_PETHY^KGLT_PETHY^Q:231-1583,H:5-448^64.46%ID^E:0^RecName: Full=Kaempferol 3-O-beta-D-galactosyltransferase
YL3_CGCTCATT-ATAGAGGC_comp1778_seq0	NA	1	BGAL5_ARATH^BGAL5_ARATH^Q:538-2550,H:61-731^71.83%ID^E:0^RecName: Full=Beta-galactosidase 5
YL3_CGCTCATT-ATAGAGGC_comp412_seq0	NA	1	RNFT2_MOUSE^RNFT2_MOUSE^Q:1847-2782,H:129-439^28.25%ID^E:1e-33^RecName: Full=RING finger and transmembrane domain-containing protein 2
YS2_GAGATTCC-ATAGAGGC_comp10747_seq0	NA	1	.
YL3_CGCTCATT-ATAGAGGC_comp10883_seq2	NA	1	DOF53_ARATH^DOF53_ARATH^Q:332-622,H:59-168^58.18%ID^E:6e-30^RecName: Full=Dof zinc finger protein DOF5.3
YL3_CGCTCATT-ATAGAGGC_comp22283_seq0	NA	1	SMT1_ARATH^SMT1_ARATH^Q:458-1018,H:149-335^78.07%ID^E:2e-106^RecName: Full=Cycloartenol-C-24-methyltransferase
OL1_CGCTCATT-CCTATCCT_comp35745_seq0	NA	1	.
YS2_GAGATTCC-ATAGAGGC_comp1818_seq4	NA	1	IDD9_ARATH^IDD9_ARATH^Q:401-1399,H:39-365^59%ID^E:2e-95^RecName: Full=Protein indeterminate-domain 9 {ECO:0000303|PubMed:16784536}
YL1_CGCTCATT-TATAGCCT_comp2414_seq3	NA	2	CCD53_ORYSJ^CCD53_ORYSJ^Q:2149-2952,H:59-328^33.09%ID^E:8e-34^RecName: Full=Cyclin-D5-3
YL3_CGCTCATT-ATAGAGGC_comp25903_seq0	NA	2	.
YL1_CGCTCATT-TATAGCCT_comp5995_seq1	NA	2	RH52A_ORYSJ^RH52A_ORYSJ^Q:509-1651,H:80-474^47.73%ID^E:1e-100^RecName: Full=DEAD-box ATP-dependent RNA helicase 52A
YS1_GAGATTCC-TATAGCCT_comp17005_seq0	NA	2	PID_ARATH^PID_ARATH^Q:220-1452,H:37-438^63.55%ID^E:3e-177^RecName: Full=Protein kinase PINOID
YS2_GAGATTCC-ATAGAGGC_comp4578_seq8	NA	2	DPOLA_ORYSJ^DPOLA_ORYSJ^Q:1194-2783,H:295-831^67.04%ID^E:0^RecName: Full=DNA polymerase alpha catalytic subunit
YS2_GAGATTCC-ATAGAGGC_comp7927_seq0	NA	2	.
YL3_CGCTCATT-ATAGAGGC_comp7058_seq2	NA	2	Y2182_ARATH^Y2182_ARATH^Q:931-3513,H:27-888^73.15%ID^E:0^.^.
YS1_GAGATTCC-TATAGCCT_comp204_seq5	NA	2	ABP20_PRUPE^ABP20_PRUPE^Q:2853-3449,H:4-202^66%ID^E:6e-77^RecName: Full=Auxin-binding protein ABP20
YS2_GAGATTCC-ATAGAGGC_comp11930_seq0	NA	2	H1_MAIZE^H1_MAIZE^Q:289-492,H:52-119^64.71%ID^E:5e-21^RecName: Full=Histone H1
YL3_CGCTCATT-ATAGAGGC_comp12994_seq0	NA	2	BGAL8_ARATH^BGAL8_ARATH^Q:171-2483,H:37-851^54.45%ID^E:0^RecName: Full=Beta-galactosidase 8
YS1_GAGATTCC-TATAGCCT_comp13633_seq0	NA	2	INVB_ARATH^INVB_ARATH^Q:188-1858,H:13-570^78.25%ID^E:0^RecName: Full=Probable alkaline/neutral invertase B {ECO:0000305}
YS1_GAGATTCC-TATAGCCT_comp1212_seq3	NA	2	GRF6_ORYSJ^GRF6_ORYSJ^Q:1117-1470,H:157-271^70.34%ID^E:1e-45^RecName: Full=Growth-regulating factor 6
YS1_GAGATTCC-TATAGCCT_comp767_seq6	NA	2	.
YL3_CGCTCATT-ATAGAGGC_comp3939_seq2	NA	2	Y4374_ARATH^Y4374_ARATH^Q:908-2677,H:29-622^61.27%ID^E:0^RecName: Full=Probable inactive receptor kinase At4g23740
OL3_CGCTCATT-GGCTCTGA_comp6787_seq2	redox1	2	NU155_ARATH^NU155_ARATH^Q:1019-5659,H:1-1464^65.74%ID^E:0^RecName: Full=Nuclear pore complex protein NUP155 {ECO:0000303|PubMed:12034489}
YS1_GAGATTCC-TATAGCCT_comp10834_seq2	NA	2	XYNF1_ASPOR^XYNF1_ASPOR^Q:1037-1792,H:42-293^29.77%ID^E:2e-28^RecName: Full=Endo-1,4-beta-xylanase F1
YS1_GAGATTCC-TATAGCCT_comp124_seq0	NA	2	LAL5_ARATH^LAL5_ARATH^Q:2436-3797,H:11-464^60.79%ID^E:8e-167^.^.

YL1_CGCTCATT-TATAGCCT_comp14048_seq1	NA	2	DPB_ARATH^DPB_ARATH^Q:259-624,H:104-226^53.23%ID^E:2e-32^RecName: Full=Transcription factor-like protein DPB
YS2_GAGATTCC-ATAGAGGC_comp48_seq0	NA	2	CIPKC_ARATH^CIPKC_ARATH^Q:329-1636,H:20-459^78.64%ID^E:0^RecName: Full=CBL-interacting serine/threonine-protein kinase 12
YL3_CGCTCATT-ATAGAGGC_comp662_seq3	NA	2	CADH6_ORYSJ^CADH6_ORYSJ^Q:95-1162,H:2-358^72.83%ID^E:1e-177^RecName: Full=Probable cinnamyl alcohol dehydrogenase 6
YS2_GAGATTCC-ATAGAGGC_comp369_seq0	NA	2	RL74_ARATH^RL74_ARATH^Q:286-1023,H:1-244^79.27%ID^E:2e-130^RecName: Full=60S ribosomal protein L7-4
YS2_GAGATTCC-ATAGAGGC_comp705_seq0	NA	2	21KD_DAUCA^21KD_DAUCA^Q:736-1218,H:28-188^55.28%ID^E:1e-58^RecName: Full=21 kDa protein
YS1_GAGATTCC-TATAGCCT_comp11479_seq0	NA	2	CMT2_ARATH^CMT2_ARATH^Q:127-813,H:726-950^53.68%ID^E:7e-116^RecName: Full=DNA (cytosine-5)-methyltransferase CMT2
YL1_CGCTCATT-TATAGCCT_comp8260_seq1	NA	2	GTOMC_ORYSJ^GTOMC_ORYSJ^Q:1-264,H:136-223^55.68%ID^E:4e-24^RecName: Full=Probable tocopherol O-methyltransferase, chloroplastic
YS1_GAGATTCC-TATAGCCT_comp17_seq1	NA	2	TET3_ARATH^TET3_ARATH^Q:1322-2176,H:1-285^68.42%ID^E:5e-125^RecName: Full=Tetraspanin-3
YS2_GAGATTCC-ATAGAGGC_comp2803_seq0	NA	2	Y3228_ARATH^Y3228_ARATH^Q:1273-1671,H:13-147^52.59%ID^E:3e-39^RecName: Full=PHD finger protein At3g20280
YL1_CGCTCATT-TATAGCCT_comp616_seq1	NA	2	CADH6_ORYSJ^CADH6_ORYSJ^Q:536-1603,H:2-358^72.83%ID^E:0^RecName: Full=Probable cinnamyl alcohol dehydrogenase 6
YL1_CGCTCATT-TATAGCCT_comp11114_seq0	NA	2	IAA17_ARATH^IAA17_ARATH^Q:84-659,H:2-219^46.33%ID^E:3e-55^RecName: Full=Auxin-responsive protein IAA17
AR2_GAGATTCC-GGCTCTGA_comp969_seq0	NA	2	H4_SOYBN^H4_SOYBN^Q:220-453,H:22-99^100%ID^E:5e-45^RecName: Full=Histone H4
YL1_CGCTCATT-TATAGCCT_comp3917_seq0	NA	2	Y5720_ARATH^Y5720_ARATH^Q:381-2264,H:30-666^52.15%ID^E:0^RecName: Full=Probable inactive receptor kinase At5g67200
YL1_CGCTCATT-TATAGCCT_comp3285_seq0	NA	2	VIL2_ARATH^VIL2_ARATH^Q:460-2259,H:2-714^36.29%ID^E:1e-94^RecName: Full=VIN3-like protein 2
YL3_CGCTCATT-ATAGAGGC_comp15965_seq1	NA	2	.
YL1_CGCTCATT-TATAGCCT_comp302_seq2	NMT	2	GTOMC_ORYSJ^GTOMC_ORYSJ^Q:1698-2549,H:75-357^48.24%ID^E:3e-88^RecName: Full=Probable tocopherol O-methyltransferase, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp12704_seq0	NA	2	NAT1_ARATH^NAT1_ARATH^Q:327-1883,H:2-520^86.71%ID^E:0^RecName: Full=Nucleobase-ascorbate transporter 1
YS2_GAGATTCC-ATAGAGGC_comp61_seq0	Vm16OMT	2	HTOMT_CATRO^HTOMT_CATRO^Q:2794-3867,H:1-355^70.28%ID^E:8e-156^RecName: Full=Tabersonine 16-O-methyltransferase
YS1_GAGATTCC-TATAGCCT_comp8503_seq3	NA	2	ROGF1_ARATH^ROGF1_ARATH^Q:764-2506,H:1-548^66.55%ID^E:0^RecName: Full=Rop guanine nucleotide exchange factor 1
YL3_CGCTCATT-ATAGAGGC_comp9464_seq1	NA	2	.
YS2_GAGATTCC-ATAGAGGC_comp5812_seq2	NA	2	LTL1_ARATH^LTL1_ARATH^Q:1690-1920,H:289-365^76.62%ID^E:8e-35^RecName: Full=GDSL esterase/lipase LTL1
YS2_GAGATTCC-ATAGAGGC_comp21866_seq0	NA	2	.
AR1_GAGATTCC-CCTATCCT_comp3145_seq1	VmPiNMT	2	GTOMC_ORYSJ^GTOMC_ORYSJ^Q:2-916,H:92-359^41.23%ID^E:4e-80^RecName: Full=Probable tocopherol O-methyltransferase, chloroplastic
AR2_GAGATTCC-GGCTCTGA_comp2837_seq1	NA	2	PPO1_IPOBA^PPO1_IPOBA^Q:341-1708,H:1-457^63.83%ID^E:0^RecName: Full=Polyphenol oxidase I, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp8982_seq1	NA	2	.
YL3_CGCTCATT-ATAGAGGC_comp2768_seq1	NA	2	.
AR1_GAGATTCC-CCTATCCT_comp1782_seq4	NA	2	KC1D_PONAB^KC1D_PONAB^Q:1149-2012,H:3-278^38.51%ID^E:6e-55^RecName: Full=Casein kinase I isoform delta
YS1_GAGATTCC-TATAGCCT_comp10782_seq1	NA	2	SMC5_ARATH^SMC5_ARATH^Q:281-3400,H:3-1042^62.21%ID^E:0^RecName: Full=Structural maintenance of chromosomes protein 5
YL1_CGCTCATT-TATAGCCT_comp17245_seq0	NA	2	.
YS2_GAGATTCC-ATAGAGGC_comp324_seq0	TDC	2	DDC_CATRO^DDC_CATRO^Q:661-2163,H:1-500^88.45%ID^E:0^RecName: Full=Aromatic-L-amino-acid decarboxylase
YL3_CGCTCATT-ATAGAGGC_comp11681_seq0	NA	2	ANXD3_ARATH^ANXD3_ARATH^Q:22-975,H:1-321^57.63%ID^E:1e-129^RecName: Full=Annexin D3
AR1_GAGATTCC-CCTATCCT_comp386_seq2	NA	2	MLO6_ARATH^MLO6_ARATH^Q:748-2376,H:5-581^59.66%ID^E:0^RecName: Full=MLO-like protein 6
YL1_CGCTCATT-TATAGCCT_comp12249_seq2	NA	2	WAXS4_ARATH^WAXS4_ARATH^Q:215-1186,H:21-337^39.2%ID^E:6e-60^RecName: Full=Probable long-chain-alcohol O-fatty-acyltransferase 4
OL1_CGCTCATT-CCTATCCT_comp6039_seq0	NA	2	.
YL1_CGCTCATT-TATAGCCT_comp202_seq1	NA	3	RUBA_PEA^RUBA_PEA^Q:1424-3199,H:1-587^86.99%ID^E:0^RecName: Full=RuBisCO large subunit-binding protein subunit alpha, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp28794_seq0	DXS3	3	DXS2_ORYSJ^DXS2_ORYSJ^Q:101-811,H:12-247^75.21%ID^E:2e-115^RecName: Full=Probable 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic
YL3_CGCTCATT-ATAGAGGC_comp22045_seq0	NA	3	.
YS2_GAGATTCC-ATAGAGGC_comp2791_seq1	NA	3	FLS2_ARATH^FLS2_ARATH^Q:110-1351,H:8-477^28.54%ID^E:2e-29^RecName: Full=LRR receptor-like serine/threonine-protein kinase FLS2
YL3_CGCTCATT-ATAGAGGC_comp3100_seq5	NA	3	PII2_ARATH^PII2_ARATH^Q:2363-3541,H:30-423^61.06%ID^E:2e-132^RecName: Full=Piriformospora indica-insensitive protein 2
AR1_GAGATTCC-CCTATCCT_comp25230_seq0	NA	3	.
YL3_CGCTCATT-ATAGAGGC_comp2756_seq3	NA	3	CELF3_BOVIN^CELF3_BOVIN^Q:830-2026,H:5-439^36.61%ID^E:1e-68^RecName: Full=CUGBP Elav-like family member 3
YL3_CGCTCATT-ATAGAGGC_comp477_seq1	IS	3	IRIS_CATRO^IRIS_CATRO^Q:199-1458,H:1-388^83.33%ID^E:0^RecName: Full=Iridoid synthase
YS2_GAGATTCC-ATAGAGGC_comp37355_seq0	NA	3	RAC1_BETVU^RAC1_BETVU^Q:900-1328,H:54-197^92.36%ID^E:5e-75^RecName: Full=Rac-like GTP-binding protein RHO1
YL3_CGCTCATT-ATAGAGGC_comp10649_seq0	NA	3	ARAK_ARATH^ARAK_ARATH^Q:617-3511,H:66-1034^72.38%ID^E:0^RecName: Full=L-arabinokinase

YS2_GAGATTCC-ATAGAGGC_comp2711_seq8	NA	3	LVSG_DICDI^LVSG_DICDI^Q:1856-2908,H:384-772^38.39%ID^E:4e-71^RecName: Full=Probable inactive serine/threonine-protein kinase lvsG
YL1_CGCTCATT-TATAGCCT_comp5166_seq0	NA	3	.
YL3_CGCTCATT-ATAGAGGC_comp3314_seq1	NA	3	Y3565_ARATH^Y3565_ARATH^Q:1642-2610,H:164-498^61.83%ID^E:1e-129^RecName: Full=Probable inactive receptor-like protein kinase At3g56050
YS1_GAGATTCC-TATAGCCT_comp1665_seq2	NA	3	BLH6_ARATH^BLH6_ARATH^Q:1000-2730,H:3-494^42.76%ID^E:3e-115^RecName: Full=BEL1-like homeodomain protein 6
YL3_CGCTCATT-ATAGAGGC_comp7757_seq0	NA	3	Y5656_ARATH^Y5656_ARATH^Q:312-2372,H:1-667^56.05%ID^E:0^RecName: Full=BTB/POZ domain-containing protein At5g66560
YL1_CGCTCATT-TATAGCCT_comp9867_seq0	NA	3	CIPK1_ORYSJ^CIPK1_ORYSJ^Q:1-276,H:148-239^76.09%ID^E:7e-75^RecName: Full=CBL-interacting protein kinase 1
AR1_GAGATTCC-CCTATCCT_comp22848_seq1	NA	3	MSI4_ARATH^MSI4_ARATH^Q:65-220,H:66-117^82.69%ID^E:3e-23^RecName: Full=WD-40 repeat-containing protein MSI4
YL1_CGCTCATT-TATAGCCT_comp7173_seq1	NA	3	RABA3_ARATH^RABA3_ARATH^Q:189-842,H:23-237^71.3%ID^E:5e-103^RecName: Full=Ras-related protein RABA3
YL1_CGCTCATT-TATAGCCT_comp6388_seq3	NA	3	PMA4_NICPL^PMA4_NICPL^Q:4708-3239,H:390-801^71.02%ID^E:0^RecName: Full=Plasma membrane ATPase 4
YL1_CGCTCATT-TATAGCCT_comp643_seq0	NA	3	RAN3_ARATH^RAN3_ARATH^Q:79-717,H:1-213^98.12%ID^E:7e-142^RecName: Full=GTP-binding nuclear protein Ran-3
YS1_GAGATTCC-TATAGCCT_comp264_seq2	NA	3	.
AR1_GAGATTCC-CCTATCCT_comp1758_seq2	NA	3	SYY_MIMIV^SYY_MIMIV^Q:868-1755,H:57-342^44.15%ID^E:1e-72^RecName: Full=Tyrosine--tRNA ligase
YS2_GAGATTCC-ATAGAGGC_comp6671_seq1	NA	3	NLTL2_ARATH^NLTL2_ARATH^Q:303-716,H:26-162^53.62%ID^E:2e-40^RecName: Full=Non-specific lipid-transfer protein-like protein At2g13820
YS1_GAGATTCC-TATAGCCT_comp14492_seq0	NA	3	RL15_PETHY^RL15_PETHY^Q:1380-1991,H:1-204^94.61%ID^E:2e-112^RecName: Full=60S ribosomal protein L15
AR2_GAGATTCC-GGCTCTGA_comp9401_seq0	NA	3	.
YL3_CGCTCATT-ATAGAGGC_comp4742_seq3	BIS1	3	BBR_ARATH^BBR_ARATH^Q:2045-2518,H:180-335^67.09%ID^E:3e-53^RecName: Full=E3 ubiquitin ligase BIG BROTHER-related
YS2_GAGATTCC-ATAGAGGC_comp5910_seq2	NA	3	VSR1_ARATH^VSR1_ARATH^Q:2339-4132,H:22-623^77.24%ID^E:0^RecName: Full=Vacuolar-sorting receptor 1
YL1_CGCTCATT-TATAGCCT_comp5093_seq2	NA	3	IAA14_ARATH^IAA14_ARATH^Q:352-843,H:4-151^48.8%ID^E:7e-32^RecName: Full=Auxin-responsive protein IAA14
AR1_GAGATTCC-CCTATCCT_comp589_seq7	G10H	3	C76B6_CATRO^C76B6_CATRO^Q:1306-2784,H:1-493^86%ID^E:0^RecName: Full=Geraniol 8-hydroxylase
AR1_GAGATTCC-CCTATCCT_comp20069_seq0	NA	3	CB4_SPIOL^CB4_SPIOL^Q:1590-1081,H:80-249^90%ID^E:1e-94^RecName: Full=Chlorophyll a-b binding protein CP24, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp10055_seq1	NA	3	.
OL3_CGCTCATT-GGCTCTGA_comp7833_seq1	NA	3	POL4_DROME^POL4_DROME^Q:611-1714,H:876-1234^25.88%ID^E:6e-29^RecName: Full=Retrovirus-related Pol polyprotein from transposon 412
YS2_GAGATTCC-ATAGAGGC_comp558_seq3	DL7GT	3	UGT8_CATRO^UGT8_CATRO^Q:834-2273,H:3-482^88.12%ID^E:0^RecName: Full=7-deoxyloganetic acid glucosyltransferase
YL1_CGCTCATT-TATAGCCT_comp20383_seq0	NA	3	.
YL1_CGCTCATT-TATAGCCT_comp5868_seq1	NA	3	VAP21_ARATH^VAP21_ARATH^Q:498-1160,H:1-220^64.29%ID^E:2e-81^RecName: Full=Vesicle-associated protein 2-1
YL1_CGCTCATT-TATAGCCT_comp7706_seq0	NA	3	.
YS1_GAGATTCC-TATAGCCT_comp5831_seq1	NA	3	.
YS1_GAGATTCC-TATAGCCT_comp10376_seq0	NA	3	.
AR1_GAGATTCC-CCTATCCT_comp4243_seq0	NA	3	.
YS2_GAGATTCC-ATAGAGGC_comp1901_seq0	NA	3	ASPG1_ARATH^ASPG1_ARATH^Q:217-1569,H:45-500^58.42%ID^E:2e-169^RecName: Full=Protein ASPARTIC PROTEASE IN GUARD CELL 1
AR1_GAGATTCC-CCTATCCT_comp3050_seq0	THAS1	4	10HGO_CATRO^10HGO_CATRO^Q:256-1335,H:1-360^86.94%ID^E:0^RecName: Full=8-hydroxygeraniol dehydrogenase
OL1_CGCTCATT-CCTATCCT_comp1163_seq6	NA	4	GATP3_SOLLC^GATP3_SOLLC^Q:1986-2429,H:370-516^80.41%ID^E:6e-74^RecName: Full=Gamma aminobutyrate transaminase 3, chloroplastic
AR1_GAGATTCC-CCTATCCT_comp8371_seq0	NA	4	HIS2_ARATH^HIS2_ARATH^Q:186-1004,H:23-279^68.5%ID^E:9e-121^RecName: Full=Histidine biosynthesis bifunctional protein hisIE, chloroplastic
AR1_GAGATTCC-CCTATCCT_comp28_seq3	NA	4	.
YL3_CGCTCATT-ATAGAGGC_comp3627_seq0	NA	4	CYC6_ARATH^CYC6_ARATH^Q:474-698,H:100-174^88%ID^E:7e-45^RecName: Full=Cytochrome c6, chloroplastic {ECO:0000303|PubMed:12037572}
YS1_GAGATTCC-TATAGCCT_comp8635_seq3	NA	4	.
YS2_GAGATTCC-ATAGAGGC_comp8318_seq1	T3O	4	C71D7_SOLCH^C71D7_SOLCH^Q:3390-3109,H:82-182^46.53%ID^E:2e-49^RecName: Full=Cytochrome P450 71D7
YS1_GAGATTCC-TATAGCCT_comp6764_seq2	V19H	4	HACL_ARATH^HACL_ARATH^Q:229-1950,H:1-572^79.62%ID^E:0^RecName: Full=2-hydroxyacyl-CoA lyase
YL3_CGCTCATT-ATAGAGGC_comp14054_seq2	NA	4	.
YS1_GAGATTCC-TATAGCCT_comp11712_seq2	NA	4	.
AR1_GAGATTCC-CCTATCCT_comp2271_seq1	NA	4	.
YL3_CGCTCATT-ATAGAGGC_comp70_seq0	8HGO	4	ADHX_ORYSI^ADHX_ORYSI^Q:149-1249,H:8-377^44.89%ID^E:1e-92^RecName: Full=Alcohol dehydrogenase class-3
YL3_CGCTCATT-ATAGAGGC_comp4424_seq2	NA	4	RGLG2_ARATH^RGLG2_ARATH^Q:675-1517,H:184-468^71.23%ID^E:5e-146^RecName: Full=E3 ubiquitin-protein ligase RGLG2
YS2_GAGATTCC-ATAGAGGC_comp6211_seq1	NA	4	KPRS4_SPIOL^KPRS4_SPIOL^Q:220-1173,H:1-318^87.74%ID^E:0^RecName: Full=Ribose-phosphate pyrophosphokinase 4

YL3_CGCTCATT-ATAGAGGC_comp25807_seq0	NA	4	.
YS2_GAGATTCC-ATAGAGGC_comp1127_seq1	NA	4	STIP1_ARATH^STIP1_ARATH^Q:1314-3857,H:1-841^53.51%ID^E:0^RecName: Full=Septin and tuftelin-interacting protein 1 homolog 1
YS1_GAGATTCC-TATAGCCT_comp72982_seq0	NA	4	.
YL1_CGCTCATT-TATAGCCT_comp13722_seq1	NA	4	PDPK2_ARATH^PDPK2_ARATH^Q:848-1456,H:32-234^91.13%ID^E:0^RecName: Full=3-phosphoinositide-dependent protein kinase 2
YS2_GAGATTCC-ATAGAGGC_comp13836_seq1	NA	4	MCM9_ARATH^MCM9_ARATH^Q:917-2809,H:12-640^75.63%ID^E:0^RecName: Full=Probable DNA helicase MCM9
YL1_CGCTCATT-TATAGCCT_comp1306_seq0	NA	4	THO4A_ARATH^THO4A_ARATH^Q:205-741,H:1-172^59.67%ID^E:5e-44^RecName: Full=THO complex subunit 4A
OL3_CGCTCATT-GGCTCTGA_comp9667_seq1	NA	4	CBSX3_ARATH^CBSX3_ARATH^Q:421-885,H:16-206^51.31%ID^E:1e-51^RecName: Full=CBS domain-containing protein CBSX3, mitochondrial
YL3_CGCTCATT-ATAGAGGC_comp14103_seq1	NA	4	.
AR1_GAGATTCC-CCTATCCT_comp4362_seq3	NA	4	RP8L2_ARATH^RP8L2_ARATH^Q:996-2741,H:345-901^31.18%ID^E:1e-68^RecName: Full=Probable disease resistance RPP8-like protein 2
OL1_CGCTCATT-CCTATCCT_comp5736_seq1	NA	5	2A5B_ARATH^2A5B_ARATH^Q:711-2147,H:1-485^70.43%ID^E:0^RecName: Full=Serine/threonine protein phosphatase 2A 57 kDa regulatory subunit B' beta isoform
OL3_CGCTCATT-GGCTCTGA_comp87_seq0	NA	5	GLNA2_VITVI^GLNA2_VITVI^Q:1-246,H:225-306^95.12%ID^E:2e-44^RecName: Full=Glutamine synthetase cytosolic isozyme 2
OL3_CGCTCATT-GGCTCTGA_comp166_seq3	NA	5	.
OL3_CGCTCATT-GGCTCTGA_comp8049_seq3	NA	5	.
OL3_CGCTCATT-GGCTCTGA_comp5964_seq3	NA	5	.
OL3_CGCTCATT-GGCTCTGA_comp3166_seq0	NA	5	C71DC_CATRO^C71DC_CATRO^Q:276-1646,H:38-492^59.04%ID^E:0^RecName: Full=Tabersonine 16-hydroxylase
OL3_CGCTCATT-GGCTCTGA_comp6046_seq2	NA	5	.
OL3_CGCTCATT-GGCTCTGA_comp3275_seq3	NA	5	LENG8_XENLA^LENG8_XENLA^Q:1359-2120,H:515-765^47.86%ID^E:4e-66^RecName: Full=Leukocyte receptor cluster member 8 homolog
OL1_CGCTCATT-CCTATCCT_comp1318_seq1	VS	5	HIDM_GLYEC^HIDM_GLYEC^Q:72-566,H:12-178^52.07%ID^E:2e-43^RecName: Full=2-hydroxyisoflavanone dehydratase
YS1_GAGATTCC-TATAGCCT_comp1103_seq2	NA	6	.
YS1_GAGATTCC-TATAGCCT_comp6148_seq1	NA	6	IDD2_ARATH^IDD2_ARATH^Q:253-624,H:1-122^76.61%ID^E:1e-62^RecName: Full=Protein indeterminate-domain 2 {ECO:0000303|PubMed:16784536}
YL1_CGCTCATT-TATAGCCT_comp20633_seq0	NA	6	Y1174_SYNP6^Y1174_SYNP6^Q:188-1219,H:2-339^55.36%ID^E:5e-100^RecName: Full=Uncharacterized lipoprotein syc1174_c
OL3_CGCTCATT-GGCTCTGA_comp4608_seq0	NA	6	ARID5_ARATH^ARID5_ARATH^Q:1224-2120,H:131-434^66.23%ID^E:2e-121^RecName: Full=AT-rich interactive domain-containing protein 5
AR2_GAGATTCC-GGCTCTGA_comp1085_seq1	NA	6	RAB7_VIGAC^RAB7_VIGAC^Q:2089-2649,H:19-206^90.96%ID^E:5e-117^RecName: Full=Ras-related protein Rab7
YS2_GAGATTCC-ATAGAGGC_comp11931_seq1	NA	6	YNB1_YEAST^YNB1_YEAST^Q:499-1446,H:51-417^31.84%ID^E:2e-47^RecName: Full=Uncharacterized protein YNL011C
OL3_CGCTCATT-GGCTCTGA_comp5315_seq0	NA	6	TOC34_PEA^TOC34_PEA^Q:691-1266,H:111-301^65.1%ID^E:6e-79^RecName: Full=Translocase of chloroplast 34
AR1_GAGATTCC-CCTATCCT_comp2374_seq0	NA	6	CLPP6_ARATH^CLPP6_ARATH^Q:485-1204,H:38-271^77.92%ID^E:1e-120^RecName: Full=ATP-dependent Clp protease proteolytic subunit 6, chloroplastic {ECO:0000303|PubMed:11299370}
OL3_CGCTCATT-GGCTCTGA_comp8459_seq0	NA	6	ARFH_ARATH^ARFH_ARATH^Q:582-2246,H:185-773^61.59%ID^E:0^RecName: Full=Auxin response factor 8
YS2_GAGATTCC-ATAGAGGC_comp3776_seq2	NA	6	CSP1_ARATH^CSP1_ARATH^Q:331-1200,H:6-299^43.56%ID^E:9e-32^RecName: Full=Cold shock protein 1
OL1_CGCTCATT-CCTATCCT_comp4311_seq2	NA	6	.
OL3_CGCTCATT-GGCTCTGA_comp9977_seq1	ORCA2	6	ERF99_ARATH^ERF99_ARATH^Q:673-903,H:91-168^73.08%ID^E:2e-33^RecName: Full=Ethylene-responsive transcription factor 13
AR1_GAGATTCC-CCTATCCT_comp3443_seq0	NA	7	F3PH_ARATH^F3PH_ARATH^Q:174-1706,H:1-508^39.96%ID^E:4e-121^RecName: Full=Flavonoid 3'-monooxygenase
AR1_GAGATTCC-CCTATCCT_comp76_seq2	NA	7	E13B_WHEAT^E13B_WHEAT^Q:367-1611,H:26-453^45.92%ID^E:9e-123^RecName: Full=Glucan endo-1,3-beta-glucosidase
AR2_GAGATTCC-GGCTCTGA_comp7052_seq0	NA	7	ANX4_FRAAN^ANX4_FRAAN^Q:82-1017,H:1-314^64.97%ID^E:9e-146^RecName: Full=Annexin-like protein RJ4
AR2_GAGATTCC-GGCTCTGA_comp35_seq9	NA	7	C72A1_CATRO^C72A1_CATRO^Q:667-2238,H:1-524^90.84%ID^E:0^RecName: Full=Secologanin synthase
YS1_GAGATTCC-TATAGCCT_comp9360_seq0	NA	7	LHTL6_ARATH^LHTL6_ARATH^Q:336-1577,H:19-432^72.95%ID^E:0^RecName: Full=Lysine histidine transporter-like 6
AR1_GAGATTCC-CCTATCCT_comp15336_seq0	NA	7	AKT1_ARATH^AKT1_ARATH^Q:195-2792,H:7-857^67.97%ID^E:0^RecName: Full=Potassium channel AKT1
AR1_GAGATTCC-CCTATCCT_comp902_seq1	NA	7	AL3H1_ARATH^AL3H1_ARATH^Q:3199-3702,H:191-358^72.02%ID^E:2e-74^RecName: Full=Aldehyde dehydrogenase family 3 member H1
AR1_GAGATTCC-CCTATCCT_comp7406_seq5	NA	7	IDD2_ARATH^IDD2_ARATH^Q:585-2090,H:1-447^52.38%ID^E:5e-124^RecName: Full=Protein indeterminate-domain 2 {ECO:0000303|PubMed:16784536}
YL1_CGCTCATT-TATAGCCT_comp695_seq0	GES	7	GERS_OCIBA^GERS_OCIBA^Q:288-1925,H:31-566^63.39%ID^E:0^RecName: Full=Geraniol synthase, chloroplastic
YL3_CGCTCATT-ATAGAGGC_comp24426_seq0	NA	8	NCPR_CATRO^NCPR_CATRO^Q:239-880,H:501-714^91.59%ID^E:2e-138^RecName: Full=NADPH--cytochrome P450 reductase {ECO:0000255|HAMAP-Rule:MF_03212}
OL3_CGCTCATT-GGCTCTGA_comp2316_seq0	MYC2	8	SYY_PARUW^SYY_PARUW^Q:1-1062,H:70-422^52.82%ID^E:1e-120^RecName: Full=Tyrosine--tRNA ligase {ECO:0000255|HAMAP-Rule:MF_02006}
YL1_CGCTCATT-TATAGCCT_comp28191_seq0	NA	8	.
YL1_CGCTCATT-TATAGCCT_comp5601_seq1	NA	8	.
YL1_CGCTCATT-TATAGCCT_comp19800_seq0	NA	8	.

YL1_CGCTCATT-TATAGCCT_comp1884_seq0	NA	8	RETOL_ARATH^RETOL_ARATH^Q:403-1920,H:32-536^59.33%ID^E:0^RecName: Full=Reticuline oxidase-like protein
AR1_GAGATTCC-CCTATCCT_comp34936_seq0	NA	8	.
YL1_CGCTCATT-TATAGCCT_comp10398_seq1	NA	8	.
OL1_CGCTCATT-CCTATCCT_comp10937_seq0	NA	8	CIPKB_ARATH^CIPKB_ARATH^Q:302-1267,H:95-421^57.88%ID^E:2e-118^RecName: Full=CBL-interacting serine/threonine-protein kinase 11
OL3_CGCTCATT-GGCTCTGA_comp34446_seq0	NA	8	.
YS2_GAGATTCC-ATAGAGGC_comp4455_seq0	NA	9	.
YS1_GAGATTCC-TATAGCCT_comp18356_seq0	NA	9	.
YL1_CGCTCATT-TATAGCCT_comp10613_seq0	ORCA5	9	EF106_ARATH^EF106_ARATH^Q:111-797,H:1-190^38.43%ID^E:8e-26^RecName: Full=Ethylene-responsive transcription factor ERF106
YS1_GAGATTCC-TATAGCCT_comp14430_seq1	NA	9	.
YS1_GAGATTCC-TATAGCCT_comp9517_seq0	NA	9	EB1C_ARATH^EB1C_ARATH^Q:515-898,H:215-328^49.62%ID^E:2e-26^RecName: Full=Microtubule-associated protein RP/EB family member 1C
YS1_GAGATTCC-TATAGCCT_comp9328_seq0	NA	9	TBL34_ARATH^TBL34_ARATH^Q:661-1365,H:171-405^53.62%ID^E:1e-107^RecName: Full=Protein trichome birefringence-like 34
YL1_CGCTCATT-TATAGCCT_comp6908_seq3	NA	9	.
YS2_GAGATTCC-ATAGAGGC_comp6201_seq1	NA	9	.
YS2_GAGATTCC-ATAGAGGC_comp92_seq2	NA	9	MYO2_ARATH^MYO2_ARATH^Q:4807-6198,H:647-1116^53.24%ID^E:5e-150^RecName: Full=Myosin-2
YS2_GAGATTCC-ATAGAGGC_comp1198_seq1	NA	9	.
YL1_CGCTCATT-TATAGCCT_comp9896_seq0	NA	9	.
YL3_CGCTCATT-ATAGAGGC_comp2341_seq2	NA	9	MTDH_FRAAN^MTDH_FRAAN^Q:2520-3485,H:35-356^67.7%ID^E:5e-139^RecName: Full=Probable mannitol dehydrogenase
YL3_CGCTCATT-ATAGAGGC_comp166_seq2	CS	9	HIDH_SOYBN^HIDH_SOYBN^Q:302-1240,H:4-318^48.91%ID^E:3e-93^RecName: Full=2-hydroxyisoflavanone dehydratase
YL1_CGCTCATT-TATAGCCT_comp27681_seq0	NA	9	.
AR1_GAGATTCC-CCTATCCT_comp1435_seq3	NA	9	RAC5_ORYSJ^RAC5_ORYSJ^Q:234-755,H:1-174^94.25%ID^E:5e-117^RecName: Full=Rac-like GTP-binding protein 5
YL1_CGCTCATT-TATAGCCT_comp11923_seq3	NA	9	ASPM_SHEEP^ASPM_SHEEP^Q:867-3644,H:702-1662^24.78%ID^E:7e-52^RecName: Full=Abnormal spindle-like microcephaly-associated protein homolog
YL3_CGCTCATT-ATAGAGGC_comp4884_seq0	NA	9	.
YS1_GAGATTCC-TATAGCCT_comp37971_seq0	NA	9	.
YS1_GAGATTCC-TATAGCCT_comp15581_seq0	NA	9	.
YL1_CGCTCATT-TATAGCCT_comp6685_seq0	NA	9	.
YL3_CGCTCATT-ATAGAGGC_comp55552_seq0	NA	9	.
OL1_CGCTCATT-CCTATCCT_comp11695_seq0	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp1556_seq7	NA	10	CASTO_LOTJA^CASTO_LOTJA^Q:660-2075,H:90-565^75.68%ID^E:0^RecName: Full=Ion channel CASTOR
AR2_GAGATTCC-GGCTCTGA_comp1876_seq1	NA	10	ARP9_ORYSI^ARP9_ORYSI^Q:6424-8193,H:1-586^62.86%ID^E:0^RecName: Full=Actin-related protein 9
OL1_CGCTCATT-CCTATCCT_comp55598_seq0	NA	10	PP2A2_ARATH^PP2A2_ARATH^Q:425-973,H:124-306^96.72%ID^E:3e-129^RecName: Full=Serine/threonine-protein phosphatase PP2A-2 catalytic subunit
OL1_CGCTCATT-CCTATCCT_comp58112_seq0	NA	10	.
YL1_CGCTCATT-TATAGCCT_comp32410_seq0	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp14565_seq0	NA	10	.
AR1_GAGATTCC-CCTATCCT_comp331_seq0	NA	10	.
YL3_CGCTCATT-ATAGAGGC_comp13576_seq1	NA	10	C3H3_ARATH^C3H3_ARATH^Q:242-1228,H:23-338^45.72%ID^E:1e-77^RecName: Full=Zinc finger CCCH domain-containing protein 3
YS2_GAGATTCC-ATAGAGGC_comp805_seq0	NA	10	GGAP1_ARATH^GGAP1_ARATH^Q:704-1879,H:1-387^76.53%ID^E:0^RecName: Full=GDP-L-galactose phosphorylase 1
YS1_GAGATTCC-TATAGCCT_comp7766_seq6	NA	10	CMBL_HUMAN^CMBL_HUMAN^Q:655-1227,H:28-217^30.96%ID^E:1e-22^RecName: Full=Carboxymethylenebutenolidase homolog
YS2_GAGATTCC-ATAGAGGC_comp38970_seq0	NA	10	.
AR1_GAGATTCC-CCTATCCT_comp1158_seq0	IDI1	10	IDI1_CAMAC^IDI1_CAMAC^Q:315-1016,H:1-234^89.32%ID^E:4e-153^RecName: Full=Isopentenyl-diphosphate Delta-isomerase I
YL3_CGCTCATT-ATAGAGGC_comp6028_seq1	NA	10	.
YL3_CGCTCATT-ATAGAGGC_comp4093_seq1	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp38763_seq0	NA	10	.
OL1_CGCTCATT-CCTATCCT_comp3640_seq0	THAS2	10	10HGO_CATRO^10HGO_CATRO^Q:196-1302,H:4-360^58.54%ID^E:8e-134^RecName: Full=8-hydroxygeraniol dehydrogenase
YS2_GAGATTCC-ATAGAGGC_comp48268_seq0	NA	10	.

YS2_GAGATTCC-ATAGAGGC_comp2162_seq0	NA	10	.
YS1_GAGATTCC-TATAGCCT_comp57747_seq0	NA	10	CID9_ARATH^CID9_ARATH^Q:5-406,H:118-251^79.85%ID^E:3e-79^RecName: Full=Polyadenylate-binding protein-interacting protein 9
YS2_GAGATTCC-ATAGAGGC_comp39036_seq0	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp12232_seq0	NA	10	.
YL3_CGCTCATT-ATAGAGGC_comp33457_seq0	NA	10	.
AR1_GAGATTCC-CCTATCCT_comp4203_seq9	NA	10	UGPA_HORVU^UGPA_HORVU^Q:2-268,H:104-192^87.64%ID^E:1e-44^RecName: Full=UTP--glucose-1-phosphate uridylyltransferase
YS2_GAGATTCC-ATAGAGGC_comp1795_seq5	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp33887_seq0	NA	10	.
YS2_GAGATTCC-ATAGAGGC_comp5239_seq2	NA	10	.
OL1_CGCTCATT-CCTATCCT_comp2934_seq3	NA	11	P2C52_ARATH^P2C52_ARATH^Q:1685-3094,H:1-467^73.68%ID^E:0^RecName: Full=Probable protein phosphatase 2C 52
YS2_GAGATTCC-ATAGAGGC_comp1245_seq0	NA	11	GAI_SOLLC^GAI_SOLLC^Q:423-2090,H:26-588^67.47%ID^E:0^RecName: Full=DELLA protein GAI
YL1_CGCTCATT-TATAGCCT_comp2035_seq0	NA	11	SDC1_ORYSJ^SDC1_ORYSJ^Q:1806-2429,H:274-481^62.02%ID^E:4e-88^RecName: Full=Serine decarboxylase 1
YS2_GAGATTCC-ATAGAGGC_comp6216_seq0	NA	11	PAS1_ARATH^PAS1_ARATH^Q:289-1812,H:1-460^65.94%ID^E:0^RecName: Full=Peptidyl-prolyl cis-trans isomerase PASTICCINO1
YS1_GAGATTCC-TATAGCCT_comp8724_seq0	NA	11	YQK1_SCHPO^YQK1_SCHPO^Q:82-999,H:3-303^33.23%ID^E:4e-35^RecName: Full=UPF0676 protein C1494.01
YS1_GAGATTCC-TATAGCCT_comp6147_seq0	NA	11	.
YL3_CGCTCATT-ATAGAGGC_comp3510_seq4	NA	11	Y1461_ARATH^Y1461_ARATH^Q:849-2270,H:88-565^69.85%ID^E:0^RecName: Full=Probable serine/threonine-protein kinase At1g54610
YL3_CGCTCATT-ATAGAGGC_comp5404_seq1	NA	11	BP73_ORYSJ^BP73_ORYSJ^Q:359-1318,H:30-372^40.23%ID^E:1e-50^RecName: Full=SAP-like protein BP-73
YL1_CGCTCATT-TATAGCCT_comp11035_seq1	NA	11	Y1491_ARATH^Y1491_ARATH^Q:480-1850,H:44-483^35.64%ID^E:6e-73^RecName: Full=Uncharacterized protein At1g04910
YL1_CGCTCATT-TATAGCCT_comp24563_seq0	NA	11	.
OL1_CGCTCATT-CCTATCCT_comp1051_seq1	NA	11	LGUL_CICAR^LGUL_CICAR^Q:878-1276,H:54-186^80.45%ID^E:3e-73^RecName: Full=Lactoylglutathione lyase
YL1_CGCTCATT-TATAGCCT_comp9848_seq1	NA	11	.
YL3_CGCTCATT-ATAGAGGC_comp2061_seq2	NA	11	ICE1_ARATH^ICE1_ARATH^Q:966-2189,H:115-494^51.91%ID^E:5e-102^RecName: Full=Transcription factor ICE1
YL3_CGCTCATT-ATAGAGGC_comp16753_seq0	NA	11	.
YS1_GAGATTCC-TATAGCCT_comp5053_seq1	NA	11	.
YL1_CGCTCATT-TATAGCCT_comp3439_seq2	NA	11	Y5133_ARATH^Y5133_ARATH^Q:1686-3032,H:10-458^60.84%ID^E:0^RecName: Full=BTB/POZ domain-containing protein At5g41330
OL1_CGCTCATT-CCTATCCT_comp2674_seq2	NA	11	SUBL_ARATH^SUBL_ARATH^Q:289-2580,H:22-755^52.8%ID^E:0^.^.
YL1_CGCTCATT-TATAGCCT_comp12940_seq0	NA	11	CDA7L_HUMAN^CDA7L_HUMAN^Q:639-935,H:343-443^50.5%ID^E:5e-26^RecName: Full=Cell division cycle-associated 7-like protein
YS1_GAGATTCC-TATAGCCT_comp15918_seq0	NA	11	.
OL3_CGCTCATT-GGCTCTGA_comp157_seq6	NA	11	COL2_ARATH^COL2_ARATH^Q:3619-4866,H:1-347^47.49%ID^E:2e-91^RecName: Full=Zinc finger protein CONSTANS-LIKE 2
YS2_GAGATTCC-ATAGAGGC_comp20_seq11	NA	11	URH2_ARATH^URH2_ARATH^Q:394-1278,H:6-282^72.88%ID^E:6e-143^RecName: Full=Probable uridine nucleosidase 2
YL1_CGCTCATT-TATAGCCT_comp255_seq0	SGD	11	SG1_RAUSE^SG1_RAUSE^Q:1583-3088,H:32-532^82.7%ID^E:0^RecName: Full=Strictosidine-O-beta-D-glucosidase
YS2_GAGATTCC-ATAGAGGC_comp528_seq2	T16H1	11	C71DC_CATRO^C71DC_CATRO^Q:2140-3633,H:1-493^74.8%ID^E:0^RecName: Full=Tabersonine 16-hydroxylase
YS1_GAGATTCC-TATAGCCT_comp5427_seq0	NA	11	.
OL3_CGCTCATT-GGCTCTGA_comp2293_seq2	NA	12	IPPK_ARATH^IPPK_ARATH^Q:1032-2057,H:104-441^56.43%ID^E:8e-113^RecName: Full=Inositol-pentakisphosphate 2-kinase
YS1_GAGATTCC-TATAGCCT_comp5896_seq1	NA	12	.
YL3_CGCTCATT-ATAGAGGC_comp3968_seq5	NA	12	PDS5B_MOUSE^PDS5B_MOUSE^Q:708-3425,H:161-1075^23.32%ID^E:1e-48^RecName: Full=Sister chromatid cohesion protein PDS5 homolog B
YL1_CGCTCATT-TATAGCCT_comp53601_seq0	NA	12	RUXG_ARATH^RUXG_ARATH^Q:150-386,H:1-79^94.94%ID^E:1e-45^RecName: Full=Probable small nuclear ribonucleoprotein G
YS2_GAGATTCC-ATAGAGGC_comp1805_seq0	NA	12	PLSP1_ARATH^PLSP1_ARATH^Q:457-1044,H:96-291^77.04%ID^E:4e-106^RecName: Full=Chloroplast processing peptidase
YL1_CGCTCATT-TATAGCCT_comp1491_seq4	DXS2	12	DXS2_ORYSJ^DXS2_ORYSJ^Q:147-2255,H:12-712^81.56%ID^E:0^RecName: Full=Probable 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic
YL1_CGCTCATT-TATAGCCT_comp20888_seq0	NA	12	.
YL1_CGCTCATT-TATAGCCT_comp3346_seq5	NA	12	PUB6_ARATH^PUB6_ARATH^Q:588-1658,H:7-381^38.32%ID^E:7e-71^RecName: Full=U-box domain-containing protein 6
YL1_CGCTCATT-TATAGCCT_comp13745_seq0	NA	12	.
YS2_GAGATTCC-ATAGAGGC_comp75153_seq0	NA	13	.
YL1_CGCTCATT-TATAGCCT_comp5567_seq1	NA	13	SIPL1_ARATH^SIPL1_ARATH^Q:240-1358,H:1-372^76.68%ID^E:0^RecName: Full=Signal peptide peptidase-like 1

YL3_CGCTCATT-ATAGAGGC_comp2760_seq1	NPF2.9	13	PTR37_ARATH^PTR37_ARATH^Q:293-1912,H:22-548^43.38%ID^E:9e-167^RecName: Full=Protein NRT1/ PTR FAMILY 2.7
YS2_GAGATTCC-ATAGAGGC_comp150_seq7	NA	13	Y1745_ARATH^Y1745_ARATH^Q:750-1634,H:8-285^35.64%ID^E:7e-32^RecName: Full=Uncharacterized protein At1g51745
OL1_CGCTCATT-CCTATCCT_comp6197_seq0	NA	13	ARFA_ARATH^ARFA_ARATH^Q:2-556,H:308-497^48.24%ID^E:8e-37^RecName: Full=Auxin response factor 1
YS2_GAGATTCC-ATAGAGGC_comp6043_seq1	NA	13	SKIP5_ARATH^SKIP5_ARATH^Q:2142-2537,H:130-273^61.81%ID^E:2e-50^RecName: Full=F-box protein SKIP5
YS2_GAGATTCC-ATAGAGGC_comp9205_seq0	NA	13	.
YS2_GAGATTCC-ATAGAGGC_comp14379_seq0	NA	13	H2AV3_ORYSJ^H2AV3_ORYSJ^Q:123-533,H:1-137^83.21%ID^E:1e-60^RecName: Full=Probable histone H2A variant 3
YL1_CGCTCATT-TATAGCCT_comp60629_seq0	NA	13	.
YS2_GAGATTCC-ATAGAGGC_comp1021_seq4	NA	13	PAS2A_ORYSJ^PAS2A_ORYSJ^Q:2278-2910,H:1-211^73.93%ID^E:2e-95^.^.
YS1_GAGATTCC-TATAGCCT_comp17484_seq0	NA	13	SILD_FORIN^SILD_FORIN^Q:1010-1288,H:9-101^61.29%ID^E:8e-23^RecName: Full=Secoisolariciresinol dehydrogenase
OL3_CGCTCATT-GGCTCTGA_comp24855_seq0	NA	14	.
YS1_GAGATTCC-TATAGCCT_comp8170_seq2	NA	14	RSH3C_ARATH^RSH3C_ARATH^Q:1477-1905,H:215-345^47.55%ID^E:1e-24^RecName: Full=Probable GTP diphosphokinase RSH3, chloroplastic
OL1_CGCTCATT-CCTATCCT_comp31_seq2	NA	14	SFH6_ARATH^SFH6_ARATH^Q:1951-2721,H:37-295^62.5%ID^E:9e-115^RecName: Full=Phosphatidylinositol/phosphatidylcholine transfer protein SFH6
OL3_CGCTCATT-GGCTCTGA_comp21501_seq0	NA	14	.
YL1_CGCTCATT-TATAGCCT_comp1669_seq5	NA	14	INO1_TOBAC^INO1_TOBAC^Q:1075-2211,H:132-510^94.2%ID^E:0^RecName: Full=Inositol-3-phosphate synthase
YL1_CGCTCATT-TATAGCCT_comp18714_seq1	NA	14	.
YL1_CGCTCATT-TATAGCCT_comp1354_seq1	GATA1	14	P2C09_ORYSJ^P2C09_ORYSJ^Q:1416-1853,H:252-413^67.28%ID^E:5e-55^RecName: Full=Probable protein phosphatase 2C 9
YL1_CGCTCATT-TATAGCCT_comp20874_seq0	NA	14	.
OL3_CGCTCATT-GGCTCTGA_comp4703_seq4	NA	14	APG3_ARATH^APG3_ARATH^Q:408-1211,H:50-317^83.58%ID^E:3e-149^RecName: Full=Peptide chain release factor APG3, chloroplastic {ECO:0000305}
YL1_CGCTCATT-TATAGCCT_comp15844_seq2	NA	14	.
AR1_GAGATTCC-CCTATCCT_comp303_seq2	NA	15	.
AR1_GAGATTCC-CCTATCCT_comp521_seq0	ZCT3	15	ZAT10_ARATH^ZAT10_ARATH^Q:199-912,H:1-223^49.8%ID^E:8e-57^RecName: Full=Zinc finger protein ZAT10
AR1_GAGATTCC-CCTATCCT_comp814_seq11	NA	15	PGLR_VITVI^PGLR_VITVI^Q:2138-3325,H:59-455^51.39%ID^E:5e-124^RecName: Full=Probable polygalacturonase
AR1_GAGATTCC-CCTATCCT_comp1653_seq0	NA	15	PLA2A_ARATH^PLA2A_ARATH^Q:172-600,H:3-145^55.17%ID^E:3e-50^RecName: Full=Phospholipase A2-alpha {ECO:0000303|PubMed:16140037}
AR1_GAGATTCC-CCTATCCT_comp2186_seq5	NA	15	B3GT7_ARATH^B3GT7_ARATH^Q:740-1732,H:88-308^52.11%ID^E:1e-79^RecName: Full=Beta-1,3-galactosyltransferase 7
AR1_GAGATTCC-CCTATCCT_comp737_seq2	ZCT1	15	ZAT12_ARATH^ZAT12_ARATH^Q:298-639,H:5-109^57.39%ID^E:2e-29^RecName: Full=Zinc finger protein ZAT12
YL3_CGCTCATT-ATAGAGGC_comp1769_seq0	NA	15	PUMP5_ARATH^PUMP5_ARATH^Q:384-1355,H:1-311^73.39%ID^E:4e-146^RecName: Full=Mitochondrial uncoupling protein 5
AR1_GAGATTCC-CCTATCCT_comp3339_seq0	NA	15	.
AR1_GAGATTCC-CCTATCCT_comp20837_seq1	NA	15	WRK33_ARATH^WRK33_ARATH^Q:514-909,H:194-328^56.74%ID^E:8e-26^RecName: Full=Probable WRKY transcription factor 33
YS1_GAGATTCC-TATAGCCT_comp11393_seq0	NA	15	.
AR2_GAGATTCC-GGCTCTGA_comp5642_seq3	NA	15	E70A1_ARATH^E70A1_ARATH^Q:306-2003,H:66-627^29.11%ID^E:4e-77^RecName: Full=Exocyst complex component EXO70A1
AR1_GAGATTCC-CCTATCCT_comp19692_seq0	NA	15	.
YL1_CGCTCATT-TATAGCCT_comp1138_seq1	NA	15	XTH2_SOYBN^XTH2_SOYBN^Q:154-936,H:30-283^66.28%ID^E:6e-117^RecName: Full=Xyloglucan endotransglucosylase/hydrolase 2 {ECO:0000250|UniProtKB:Q38857}
YL1_CGCTCATT-TATAGCCT_comp2433_seq0	NA	15	CML31_ORYSJ^CML31_ORYSJ^Q:354-752,H:15-150^49.26%ID^E:5e-32^RecName: Full=Probable calcium-binding protein CML31
AR1_GAGATTCC-CCTATCCT_comp15923_seq0	NA	15	.
AR1_GAGATTCC-CCTATCCT_comp15349_seq0	NA	15	.
AR2_GAGATTCC-GGCTCTGA_comp25525_seq0	NA	15	.
YL1_CGCTCATT-TATAGCCT_comp1591_seq0	NA	15	RAP24_ARATH^RAP24_ARATH^Q:953-1651,H:76-329^56.03%ID^E:7e-60^RecName: Full=Ethylene-responsive transcription factor RAP2-4
AR1_GAGATTCC-CCTATCCT_comp3391_seq0	NA	15	.
AR1_GAGATTCC-CCTATCCT_comp19574_seq0	NA	15	.
YL1_CGCTCATT-TATAGCCT_comp8028_seq1	NA	15	NAC81_ARATH^NAC81_ARATH^Q:626-901,H:54-147^90.43%ID^E:2e-39^RecName: Full=Protein ATAF2
AR1_GAGATTCC-CCTATCCT_comp21765_seq0	NA	15	.
AR2_GAGATTCC-GGCTCTGA_comp3473_seq2	NA	15	WRK40_ARATH^WRK40_ARATH^Q:1211-1579,H:183-302^48.82%ID^E:3e-24^RecName: Full=Probable WRKY transcription factor 40
YL1_CGCTCATT-TATAGCCT_comp20940_seq0	NA	15	.
AR2_GAGATTCC-GGCTCTGA_comp1098_seq4	NA	15	RTNLQ_ARATH^RTNLQ_ARATH^Q:222-1385,H:4-373^50%ID^E:1e-91^RecName: Full=Reticulon-like protein B17

AR1_GAGATTCC-CCTATCCT_comp26846_seq1	NA	15	.
YL1_CGCTCATT-TATAGCCT_comp34042_seq0	NA	15	.
OL3_CGCTCATT-GGCTCTGA_comp30250_seq0	NA	15	RS13_SOYBN^RS13_SOYBN^Q:100-519,H:1-140^92.86%ID^E:5e-86^RecName: Full=40S ribosomal protein S13
AR2_GAGATTCC-GGCTCTGA_comp2596_seq1	NA	15	RING1_GOSHI^RING1_GOSHI^Q:682-1068,H:163-287^46.15%ID^E:1e-31^RecName: Full=E3 ubiquitin-protein ligase RING1
AR1_GAGATTCC-CCTATCCT_comp21983_seq0	NA	15	.
OL3_CGCTCATT-GGCTCTGA_comp62954_seq0	NA	15	ARFF_ORYSJ^ARFF_ORYSJ^Q:192-365,H:164-221^89.66%ID^E:3e-51^RecName: Full=Auxin response factor 6
YL1_CGCTCATT-TATAGCCT_comp2206_seq7	NA	15	WRK40_ARATH^WRK40_ARATH^Q:1321-1728,H:165-302^44.59%ID^E:1e-21^RecName: Full=Probable WRKY transcription factor 40
AR2_GAGATTCC-GGCTCTGA_comp656_seq7	NA	15	COBRA_ARATH^COBRA_ARATH^Q:469-1668,H:35-435^83.79%ID^E:0^RecName: Full=Protein COBRA
AR1_GAGATTCC-CCTATCCT_comp2897_seq1	NA	15	.
AR2_GAGATTCC-GGCTCTGA_comp630_seq3	CrERF5	15	ERF5_NICSY^ERF5_NICSY^Q:157-1017,H:1-282^46.28%ID^E:6e-47^RecName: Full=Ethylene-responsive transcription factor 5
YL1_CGCTCATT-TATAGCCT_comp1014_seq1	NA	16	QCR72_ARATH^QCR72_ARATH^Q:416-775,H:3-122^73.33%ID^E:2e-53^RecName: Full=Cytochrome b-c1 complex subunit 7-2
AR2_GAGATTCC-GGCTCTGA_comp36070_seq0	NA	16	VAP12_ARATH^VAP12_ARATH^Q:1450-2037,H:40-221^62.63%ID^E:1e-71^RecName: Full=Vesicle-associated protein 1-2
AR2_GAGATTCC-GGCTCTGA_comp10915_seq2	NA	16	.
AR1_GAGATTCC-CCTATCCT_comp4311_seq2	NA	16	.
YL1_CGCTCATT-TATAGCCT_comp32_seq1	GO	16	C7BL3_CICIN^C7BL3_CICIN^Q:361-1737,H:37-488^38.7%ID^E:9e-111^RecName: Full=Costunolide synthase
YS2_GAGATTCC-ATAGAGGC_comp7642_seq3	NA	16	HCS2_ARATH^HCS2_ARATH^Q:280-1245,H:9-329^65.02%ID^E:1e-143^RecName: Full=Biotin--protein ligase 2
OL1_CGCTCATT-CCTATCCT_comp6428_seq0	NA	16	M2K9_ARATH^M2K9_ARATH^Q:288-1262,H:1-308^60.43%ID^E:2e-114^RecName: Full=Mitogen-activated protein kinase kinase 9
YL1_CGCTCATT-TATAGCCT_comp32588_seq0	NA	16	.
OL3_CGCTCATT-GGCTCTGA_comp1418_seq0	T19H	17	C71A2_SOLME^C71A2_SOLME^Q:587-1945,H:46-503^50.33%ID^E:6e-151^RecName: Full=Cytochrome P450 71A2
OL3_CGCTCATT-GGCTCTGA_comp132_seq7	NA	17	FRO6_ARATH^FRO6_ARATH^Q:3030-4742,H:189-738^57.9%ID^E:0^RecName: Full=Ferric reduction oxidase 6
OL3_CGCTCATT-GGCTCTGA_comp15162_seq0	NA	17	.
YL1_CGCTCATT-TATAGCCT_comp12179_seq1	NA	17	BOR4_ARATH^BOR4_ARATH^Q:168-2138,H:6-669^69.66%ID^E:0^RecName: Full=Boron transporter 4
OL3_CGCTCATT-GGCTCTGA_comp7365_seq1	NA	17	RNHX1_ARATH^RNHX1_ARATH^Q:783-2540,H:1-584^25.62%ID^E:5e-41^RecName: Full=Putative ribonuclease H protein At1g65750
OL3_CGCTCATT-GGCTCTGA_comp1249_seq0	NA	17	.
OL1_CGCTCATT-CCTATCCT_comp119_seq1	NA	17	.
OL1_CGCTCATT-CCTATCCT_comp368_seq0	NA	18	METE_CATRO^METE_CATRO^Q:156-2450,H:1-765^95.16%ID^E:0^RecName: Full=5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
AR2_GAGATTCC-GGCTCTGA_comp3587_seq0	NA	18	IDH1_ARATH^IDH1_ARATH^Q:613-1503,H:67-367^87.71%ID^E:8e-171^RecName: Full=Isocitrate dehydrogenase [NAD] regulatory subunit 1, mitochondrial
YL3_CGCTCATT-ATAGAGGC_comp5981_seq5	NA	18	APY2_ARATH^APY2_ARATH^Q:2126-2938,H:201-471^71.59%ID^E:1e-120^RecName: Full=Apyrase 2
YS2_GAGATTCC-ATAGAGGC_comp1097_seq5	NA	18	.
AR1_GAGATTCC-CCTATCCT_comp6659_seq0	NA	18	.
AR2_GAGATTCC-GGCTCTGA_comp7061_seq1	NA	18	PXL1_ARATH^PXL1_ARATH^Q:167-3247,H:1-1029^63.89%ID^E:0^RecName: Full=Leucine-rich repeat receptor-like protein kinase PXL1
YL1_CGCTCATT-TATAGCCT_comp13796_seq1	NA	18	S61G2_ARATH^S61G2_ARATH^Q:3-206,H:2-69^89.71%ID^E:9e-23^RecName: Full=Protein transport protein Sec61 subunit gamma-2
AR1_GAGATTCC-CCTATCCT_comp2488_seq0	NA	18	ACS_ARATH^ACS_ARATH^Q:1426-1989,H:540-727^87.77%ID^E:8e-94^RecName: Full=Acetyl-coenzyme A synthetase, chloroplastic/glyoxysomal
AR2_GAGATTCC-GGCTCTGA_comp5665_seq0	ZCT2	18	PI3K_ARATH^PI3K_ARATH^Q:1-309,H:712-814^95.15%ID^E:4e-57^RecName: Full=Phosphatidylinositol 3-kinase VPS34
AR1_GAGATTCC-CCTATCCT_comp7698_seq2	NA	18	BIG1_ARATH^BIG1_ARATH^Q:1905-7001,H:2-1685^66.3%ID^E:0^RecName: Full=Brefeldin A-inhibited guanine nucleotide-exchange protein 1
AR1_GAGATTCC-CCTATCCT_comp2717_seq1	NA	18	TMCO1_RAT^TMCO1_RAT^Q:195-743,H:5-184^44.15%ID^E:5e-40^RecName: Full=Transmembrane and coiled-coil domains protein 1
YL3_CGCTCATT-ATAGAGGC_comp3830_seq1	NA	18	.
AR1_GAGATTCC-CCTATCCT_comp8039_seq0	GPPS	18	SPS3_ARATH^SPS3_ARATH^Q:300-1559,H:1-422^66.59%ID^E:0^RecName: Full=Solanesyl diphosphate synthase 3, chloroplastic/mitochondrial
AR1_GAGATTCC-CCTATCCT_comp17606_seq0	NA	18	.
YS2_GAGATTCC-ATAGAGGC_comp287_seq1	HDS	18	ISPG_ARATH^ISPG_ARATH^Q:239-2449,H:1-737^86.02%ID^E:0^RecName: Full=4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (ferredoxin), chloroplastic
AR2_GAGATTCC-GGCTCTGA_comp1042_seq4	NA	18	.
AR1_GAGATTCC-CCTATCCT_comp1722_seq0	NA	18	THCAS_CANSA^THCAS_CANSA^Q:162-1625,H:46-540^45.96%ID^E:2e-130^RecName: Full=Tetrahydrocannabinolic acid synthase
OL3_CGCTCATT-GGCTCTGA_comp4392_seq1	NA	18	CYC_ABUTH^CYC_ABUTH^Q:140-472,H:1-111^94.59%ID^E:2e-70^RecName: Full=Cytochrome c
YL1_CGCTCATT-TATAGCCT_comp6904_seq2	NA	18	INVA_VICFA^INVA_VICFA^Q:6713-8377,H:71-630^70%ID^E:0^RecName: Full=Acid beta-fructofuranosidase

YL3_CGCTCATT-ATAGAGGC_comp826_seq0	NA	18	14334_SOLLC^14334_SOLLC^Q:652-1428,H:4-260^79.15%ID^E:9e-140^RecName: Full=14-3-3 protein 4
YL3_CGCTCATT-ATAGAGGC_comp116_seq0	NA	18	RL9_PEA^RL9_PEA^Q:447-815,H:1-123^82.11%ID^E:7e-67^RecName: Full=60S ribosomal protein L9
YL3_CGCTCATT-ATAGAGGC_comp3473_seq0	NA	18	METK1_CATRO^METK1_CATRO^Q:99-1277,H:1-393^96.95%ID^E:0^RecName: Full=S-adenosylmethionine synthase 1
AR1_GAGATTCC-CCTATCCT_comp1014_seq2	NA	18	NOTUM_MOUSE^NOTUM_MOUSE^Q:2970-3719,H:106-368^29.7%ID^E:3e-23^RecName: Full=Palmitoleoyl-protein carboxylesterase NOTUM {ECO:0000250|UniProtKB:Q6P988}
YL3_CGCTCATT-ATAGAGGC_comp56582_seq0	NA	18	PKHA8_RAT^PKHA8_RAT^Q:531-965,H:327-471^34.93%ID^E:6e-21^RecName: Full=Pleckstrin homology domain-containing family A member 8
AR1_GAGATTCC-CCTATCCT_comp46_seq0	NA	18	SAHH_CATRO^SAHH_CATRO^Q:259-1713,H:1-485^94.85%ID^E:0^RecName: Full=Adenosylhomocysteinase
AR1_GAGATTCC-CCTATCCT_comp12098_seq0	NA	18	UBP1_ARATH^UBP1_ARATH^Q:934-1974,H:785-1083^40.17%ID^E:2e-56^RecName: Full=Ubiquitin carboxyl-terminal hydrolase 1
OL1_CGCTCATT-CCTATCCT_comp1720_seq1	NA	19	FLS_PETHY^FLS_PETHY^Q:113-1117,H:13-348^75%ID^E:2e-157^RecName: Full=Flavonol synthase/flavanone 3-hydroxylase
YS2_GAGATTCC-ATAGAGGC_comp9582_seq0	NA	19	HYES_HUMAN^HYES_HUMAN^Q:110-1027,H:237-544^33.02%ID^E:4e-36^RecName: Full=Bifunctional epoxide hydrolase 2
OL1_CGCTCATT-CCTATCCT_comp520_seq1	IO	19	C76A2_SOLME^C76A2_SOLME^Q:383-1777,H:43-503^50.97%ID^E:6e-168^RecName: Full=Cytochrome P450 76A2
YS1_GAGATTCC-TATAGCCT_comp5894_seq4	NA	19	PERK1_ARATH^PERK1_ARATH^Q:1022-2218,H:252-651^79.25%ID^E:0^RecName: Full=Proline-rich receptor-like protein kinase PERK1
AR1_GAGATTCC-CCTATCCT_comp37384_seq0	NA	19	.
YL1_CGCTCATT-TATAGCCT_comp24385_seq0	NA	19	TRBP2_ARATH^TRBP2_ARATH^Q:3-227,H:165-239^98.67%ID^E:2e-39^RecName: Full=Tryptophan synthase beta chain 2, chloroplastic
YL3_CGCTCATT-ATAGAGGC_comp17087_seq0	NA	19	.
YS2_GAGATTCC-ATAGAGGC_comp5954_seq0	NA	20	E70B1_ARATH^E70B1_ARATH^Q:239-2152,H:1-620^64.75%ID^E:0^RecName: Full=Exocyst complex component EXO70B1
YL3_CGCTCATT-ATAGAGGC_comp11008_seq1	NA	20	.
OL3_CGCTCATT-GGCTCTGA_comp13259_seq1	NA	20	CLPX_SPHWW^CLPX_SPHWW^Q:642-1646,H:65-404^54.55%ID^E:4e-118^RecName: Full=ATP-dependent Clp protease ATP-binding subunit ClpX {ECO:0000255|HAMAP-Rule:MF_00175}
YL3_CGCTCATT-ATAGAGGC_comp32910_seq0	NA	20	.
YS2_GAGATTCC-ATAGAGGC_comp6479_seq1	NA	20	ADT1_GOSHI^ADT1_GOSHI^Q:155-1318,H:1-386^83.76%ID^E:0^RecName: Full=ADP,ATP carrier protein 1, mitochondrial
AR1_GAGATTCC-CCTATCCT_comp3108_seq0	T3R	20	10HGO_CATRO^10HGO_CATRO^Q:164-1228,H:2-360^63.51%ID^E:1e-145^RecName: Full=8-hydroxygeraniol dehydrogenase
OL1_CGCTCATT-CCTATCCT_comp209_seq10	NA	20	.
YS1_GAGATTCC-TATAGCCT_comp6840_seq0	NA	20	.
YL3_CGCTCATT-ATAGAGGC_comp30598_seq0	NA	20	.
AR2_GAGATTCC-GGCTCTGA_comp29260_seq0	NA	21	.
AR1_GAGATTCC-CCTATCCT_comp14477_seq0	ORCA3	21	ERF99_ARATH^ERF99_ARATH^Q:401-604,H:70-136^72.06%ID^E:6e-25^RecName: Full=Ethylene-responsive transcription factor 13
AR1_GAGATTCC-CCTATCCT_comp6627_seq1	NA	21	.
AR1_GAGATTCC-CCTATCCT_comp18379_seq1	NA	21	.
AR1_GAGATTCC-CCTATCCT_comp2945_seq0	NA	21	RA51B_ARATH^RA51B_ARATH^Q:6971-8011,H:26-370^66.57%ID^E:4e-122^RecName: Full=DNA repair protein RAD51 homolog 2
YS2_GAGATTCC-ATAGAGGC_comp9578_seq1	NA	21	Y4374_ARATH^Y4374_ARATH^Q:320-2167,H:18-635^56.05%ID^E:0^RecName: Full=Probable inactive receptor kinase At4g23740
OL3_CGCTCATT-GGCTCTGA_comp25748_seq0	NA	22	.
YS1_GAGATTCC-TATAGCCT_comp40843_seq0	NA	22	.
YS1_GAGATTCC-TATAGCCT_comp8051_seq0	NA	22	.
OL3_CGCTCATT-GGCTCTGA_comp21659_seq0	NA	22	.
AR2_GAGATTCC-GGCTCTGA_comp12782_seq0	NA	22	.
OL3_CGCTCATT-GGCTCTGA_comp34622_seq0	NA	22	.
OL3_CGCTCATT-GGCTCTGA_comp2285_seq1	NPF2.1	22	PTR21_ARATH^PTR21_ARATH^Q:684-2408,H:44-609^54.43%ID^E:0^RecName: Full=Protein NRT1/ PTR FAMILY 2.13
YS1_GAGATTCC-TATAGCCT_comp64350_seq0	NA	22	.
YL1_CGCTCATT-TATAGCCT_comp16103_seq0	NA	22	PEX13_ARATH^PEX13_ARATH^Q:451-735,H:202-304^66.02%ID^E:1e-27^RecName: Full=Peroxisomal membrane protein 13
YS2_GAGATTCC-ATAGAGGC_comp11235_seq0	NA	23	CX32_ARATH^CX32_ARATH^Q:640-1569,H:63-369^69.03%ID^E:2e-150^RecName: Full=Probable serine/threonine-protein kinase Cx32, chloroplastic
OL3_CGCTCATT-GGCTCTGA_comp18596_seq0	NA	23	.
OL1_CGCTCATT-CCTATCCT_comp22262_seq0	NA	23	BPS1_ARATH^BPS1_ARATH^Q:994-1986,H:1-326^56.12%ID^E:4e-123^RecName: Full=Protein BPS1, chloroplastic
OL3_CGCTCATT-GGCTCTGA_comp2267_seq1	VmTPT2	23	AB32G_ARATH^AB32G_ARATH^Q:1321-5580,H:1-1418^70.34%ID^E:0^RecName: Full=ABC transporter G family member 32 {ECO:0000303|PubMed:18299247}
YL1_CGCTCATT-TATAGCCT_comp12043_seq2	NA	23	BCS1_SCHPO^BCS1_SCHPO^Q:1059-1733,H:195-423^29.05%ID^E:1e-24^RecName: Full=Probable mitochondrial chaperone bcs1
OL1_CGCTCATT-CCTATCCT_comp671_seq2	NA	23	SEC8_ARATH^SEC8_ARATH^Q:2372-5161,H:143-1053^68.17%ID^E:0^RecName: Full=Exocyst complex component SEC8

OL1_CGCTCATT-CCTATCCT_comp13536_seq1	NA	23	.
YS1_GAGATTCC-TATAGCCT_comp17664_seq1	NA	23	P2C51_ARATH^P2C51_ARATH^Q:134-1405,H:30-447^54.55%ID^E:1e-139^RecName: Full=Probable protein phosphatase 2C 51
OL3_CGCTCATT-GGCTCTGA_comp2330_seq3	NA	24	.
YL3_CGCTCATT-ATAGAGGC_comp15500_seq0	NA	24	Y1680_ARATH^Y1680_ARATH^Q:2339-4210,H:30-648^42.17%ID^E:9e-86^RecName: Full=Probable leucine-rich repeat receptor-like protein kinase At1g68400
YL1_CGCTCATT-TATAGCCT_comp8450_seq0	NA	24	YERG_SCHPO^YERG_SCHPO^Q:37-378,H:204-320^42.74%ID^E:8e-23^RecName: Full=Uncharacterized RING finger protein C2F3.16
YS1_GAGATTCC-TATAGCCT_comp15900_seq1	NA	24	SIGD_ARATH^SIGD_ARATH^Q:957-1427,H:258-414^54.78%ID^E:9e-82^RecName: Full=RNA polymerase sigma factor sigD, chloroplastic
OL3_CGCTCATT-GGCTCTGA_comp3385_seq1	NA	24	.
YS2_GAGATTCC-ATAGAGGC_comp1361_seq4	NA	24	CYB5E_ARATH^CYB5E_ARATH^Q:2054-2449,H:2-133^81.82%ID^E:2e-70^RecName: Full=Cytochrome b5 isoform E {ECO:0000303|PubMed:19054355}
OL3_CGCTCATT-GGCTCTGA_comp8789_seq0	NPF2.4	24	PTR53_ARATH^PTR53_ARATH^Q:121-1830,H:27-597^57.62%ID^E:0^RecName: Full=Protein NRT1/ PTR FAMILY 2.11
OL3_CGCTCATT-GGCTCTGA_comp35783_seq0	NA	24	.
OL3_CGCTCATT-GGCTCTGA_comp19373_seq0	NA	24	.
AR2_GAGATTCC-GGCTCTGA_comp2182_seq0	NA	25	XCP1_ARATH^XCP1_ARATH^Q:155-1150,H:26-353^72.59%ID^E:0^RecName: Full=Xylem cysteine proteinase 1
YS1_GAGATTCC-TATAGCCT_comp17355_seq0	NA	25	.
YL3_CGCTCATT-ATAGAGGC_comp25012_seq1	NA	25	CSLCC_ARATH^CSLCC_ARATH^Q:1254-3374,H:1-699^78.22%ID^E:0^RecName: Full=Probable xyloglucan glycosyltransferase 12
YS2_GAGATTCC-ATAGAGGC_comp3345_seq1	MATE2	25	ALIS3_ARATH^ALIS3_ARATH^Q:2756-3742,H:22-349^75.99%ID^E:8e-171^RecName: Full=ALA-interacting subunit 3
AR2_GAGATTCC-GGCTCTGA_comp901_seq0	NA	25	.
AR1_GAGATTCC-CCTATCCT_comp2521_seq4	NA	25	UPL3_ARATH^UPL3_ARATH^Q:848-6109,H:151-1888^71.65%ID^E:0^RecName: Full=E3 ubiquitin-protein ligase UPL3
YS1_GAGATTCC-TATAGCCT_comp1496_seq2	CMK	26	ISPE_SOLLC^ISPE_SOLLC^Q:362-1471,H:31-399^75.95%ID^E:2e-175^RecName: Full=4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, chloroplastic/chromoplastic
AR2_GAGATTCC-GGCTCTGA_comp7989_seq1	NA	26	.
AR1_GAGATTCC-CCTATCCT_comp4372_seq0	NA	26	.
YS2_GAGATTCC-ATAGAGGC_comp147_seq0	NA	26	RS33_ARATH^RS33_ARATH^Q:169-834,H:1-222^87.84%ID^E:3e-142^RecName: Full=40S ribosomal protein S3-3
YL1_CGCTCATT-TATAGCCT_comp29003_seq0	NA	26	.
AR1_GAGATTCC-CCTATCCT_comp1690_seq3	NA	26	PMT2_ARATH^PMT2_ARATH^Q:374-2215,H:1-616^76.45%ID^E:0^RecName: Full=Probable methyltransferase PMT2
YS2_GAGATTCC-ATAGAGGC_comp21925_seq0	NA	26	.
AR1_GAGATTCC-CCTATCCT_comp3686_seq1	NA	27	.
AR1_GAGATTCC-CCTATCCT_comp15754_seq0	NA	27	.
AR1_GAGATTCC-CCTATCCT_comp10689_seq2	DAT	27	DAT_CATRO^DAT_CATRO^Q:1-1251,H:5-437^54.71%ID^E:5e-150^RecName: Full=Deacetylvindoline O-acetyltransferase
AR2_GAGATTCC-GGCTCTGA_comp216_seq0	NA	27	.
AR2_GAGATTCC-GGCTCTGA_comp22743_seq0	NA	27	.
OL3_CGCTCATT-GGCTCTGA_comp5631_seq2	GBF2	28	HBP1A_WHEAT^HBP1A_WHEAT^Q:439-1467,H:18-319^45.3%ID^E:4e-54^RecName: Full=Transcription factor HBP-1a
AR2_GAGATTCC-GGCTCTGA_comp2924_seq0	NA	28	.
AR1_GAGATTCC-CCTATCCT_comp4693_seq1	NA	28	.
YL3_CGCTCATT-ATAGAGGC_comp15371_seq1	NA	28	RDR5_ARATH^RDR5_ARATH^Q:546-3176,H:114-975^48.45%ID^E:0^RecName: Full=Probable RNA-dependent RNA polymerase 5
YL3_CGCTCATT-ATAGAGGC_comp4673_seq1	NA	28	NSA2_HUMAN^NSA2_HUMAN^Q:2-265,H:149-236^78.41%ID^E:1e-42^RecName: Full=Ribosome biogenesis protein NSA2 homolog
AR2_GAGATTCC-GGCTCTGA_comp9703_seq0	NA	28	INVB_DAUCA^INVB_DAUCA^Q:1146-3059,H:4-656^67.48%ID^E:0^RecName: Full=Beta-fructofuranosidase, soluble isoenzyme I
YS2_GAGATTCC-ATAGAGGC_comp6465_seq1	NA	29	Y283_BACHD^Y283_BACHD^Q:829-1725,H:1-265^39.33%ID^E:1e-58^RecName: Full=Uncharacterized isomerase BH0283
YL3_CGCTCATT-ATAGAGGC_comp37282_seq0	NA	29	.
AR1_GAGATTCC-CCTATCCT_comp12730_seq0	NA	29	.
YL3_CGCTCATT-ATAGAGGC_comp1199_seq2	NA	29	GSTX4_TOBAC^GSTX4_TOBAC^Q:401-1045,H:5-220^68.52%ID^E:6e-101^RecName: Full=Probable glutathione S-transferase
YL3_CGCTCATT-ATAGAGGC_comp3815_seq0	NA	29	CSK2C_ARATH^CSK2C_ARATH^Q:57-521,H:96-250^80%ID^E:2e-89^RecName: Full=Casein kinase II subunit beta'
YS2_GAGATTCC-ATAGAGGC_comp88_seq3	LAMT	29	MT878_ARATH^MT878_ARATH^Q:389-1450,H:3-361^41.1%ID^E:4e-79^RecName: Full=Probable S-adenosylmethionine-dependent methyltransferase At5g38780
YS1_GAGATTCC-TATAGCCT_comp3264_seq4	NA	29	LOR12_ARATH^LOR12_ARATH^Q:1818-2432,H:9-205^50%ID^E:5e-54^RecName: Full=Protein LURP-one-related 12
AR2_GAGATTCC-GGCTCTGA_comp32417_seq0	NA	29	.
AR1_GAGATTCC-CCTATCCT_comp19025_seq0	NA	29	C2D61_ARATH^C2D61_ARATH^Q:2-235,H:285-362^79.49%ID^E:4e-34^RecName: Full=C2 domain-containing protein At1g53590

AR1_GAGATTCC-CCTATCCT_comp4621_seq0	NA	29	BB_ARATH^BB_ARATH^Q:1423-1734,H:138-241^66.67%ID^E:5e-39^RecName: Full=E3 ubiquitin ligase BIG BROTHER
AR1_GAGATTCC-CCTATCCT_comp2309_seq1	NA	29	HHP2_ARATH^HHP2_ARATH^Q:519-1337,H:75-356^60.07%ID^E:4e-79^RecName: Full=Heptahelical transmembrane protein 2
YS1_GAGATTCC-TATAGCCT_comp9272_seq0	NA	29	KPYC_SOLTU^KPYC_SOLTU^Q:840-1322,H:347-506^41.61%ID^E:6e-31^RecName: Full=Pyruvate kinase, cytosolic isozyme
YS1_GAGATTCC-TATAGCCT_comp439_seq2	HDR	30	ISPH_ARATH^ISPH_ARATH^Q:147-1547,H:1-466^78.09%ID^E:0^RecName: Full=4-hydroxy-3-methylbut-2-enyl diphosphate reductase, chloroplastic
OL1_CGCTCATT-CCTATCCT_comp161_seq5	NA	30	FRO7_ARATH^FRO7_ARATH^Q:3155-5278,H:32-747^59.58%ID^E:0^RecName: Full=Ferric reduction oxidase 7, chloroplastic
OL1_CGCTCATT-CCTATCCT_comp991_seq7	NA	30	.
OL1_CGCTCATT-CCTATCCT_comp5868_seq2	NA	30	MOC2B_ARATH^MOC2B_ARATH^Q:1951-2445,H:18-183^66.87%ID^E:7e-57^RecName: Full=Molybdopterin synthase catalytic subunit {ECO:0000255|HAMAP-Rule:MF_03052}
YS2_GAGATTCC-ATAGAGGC_comp6690_seq0	NA	30	PERR_RAUSE^PERR_RAUSE^Q:232-1107,H:1-289^74.66%ID^E:2e-154^RecName: Full=Perakine reductase
OL1_CGCTCATT-CCTATCCT_comp9509_seq1	NA	30	Y4885_ARATH^Y4885_ARATH^Q:309-2498,H:312-1041^42.57%ID^E:2e-172^RecName: Full=Probable LRR receptor-like serine/threonine-protein kinase At4g08850
YL1_CGCTCATT-TATAGCCT_comp615_seq1	NA	30	STR4_ARATH^STR4_ARATH^Q:734-1729,H:1-321^62.46%ID^E:5e-87^RecName: Full=Rhodanese-like domain-containing protein 4, chloroplastic
YS2_GAGATTCC-ATAGAGGC_comp9283_seq0	NA	30	Y1480_ARATH^Y1480_ARATH^Q:312-1265,H:55-357^39.69%ID^E:3e-59^RecName: Full=Uncharacterized protein At1g18480
YL1_CGCTCATT-TATAGCCT_comp4903_seq0	NA	30	NIFU1_ARATH^NIFU1_ARATH^Q:212-703,H:66-230^79.39%ID^E:3e-63^RecName: Full=NifU-like protein 1, chloroplastic
OL1_CGCTCATT-CCTATCCT_comp7103_seq2	NA	30	RH41_ARATH^RH41_ARATH^Q:768-2225,H:12-490^59.8%ID^E:0^RecName: Full=DEAD-box ATP-dependent RNA helicase 41
YL3_CGCTCATT-ATAGAGGC_comp1376_seq3	NA	30	RVE8_ARATH^RVE8_ARATH^Q:203-565,H:31-151^82.64%ID^E:6e-47^RecName: Full=Protein REVEILLE 8
YL1_CGCTCATT-TATAGCCT_comp3675_seq3	NA	30	SIGE_ARATH^SIGE_ARATH^Q:721-2280,H:1-516^62.93%ID^E:0^RecName: Full=RNA polymerase sigma factor sigE, chloroplastic/mitochondrial
YL1_CGCTCATT-TATAGCCT_comp12033_seq1	NA	31	C90A1_ARATH^C90A1_ARATH^Q:285-1586,H:41-468^75.86%ID^E:0^RecName: Full=Cytochrome P450 90A1
YS2_GAGATTCC-ATAGAGGC_comp2127_seq1	DXS1	31	PWD_ARATH^PWD_ARATH^Q:2912-6307,H:48-1196^63.9%ID^E:0^RecName: Full=Phosphoglucan, water dikinase, chloroplastic
YL3_CGCTCATT-ATAGAGGC_comp4506_seq1	NA	31	.
YS2_GAGATTCC-ATAGAGGC_comp11659_seq0	NA	31	.
YS2_GAGATTCC-ATAGAGGC_comp8026_seq0	NA	31	.
YL1_CGCTCATT-TATAGCCT_comp11163_seq2	NA	31	.
YL3_CGCTCATT-ATAGAGGC_comp38593_seq0	NA	31	SYE_ARATH^SYE_ARATH^Q:709-1746,H:223-568^76.37%ID^E:0^.^.
AR2_GAGATTCC-GGCTCTGA_comp894_seq1	NA	32	.
AR1_GAGATTCC-CCTATCCT_comp8273_seq2	NA	32	S38A7_DANRE^S38A7_DANRE^Q:251-1336,H:56-409^25.33%ID^E:2e-25^RecName: Full=Putative sodium-coupled neutral amino acid transporter 7
YS1_GAGATTCC-TATAGCCT_comp13751_seq2	TS	32	LAX3_MEDTR^LAX3_MEDTR^Q:2-1273,H:42-465^89.15%ID^E:0^RecName: Full=Auxin transporter-like protein 3
AR2_GAGATTCC-GGCTCTGA_comp3252_seq1	NA	32	CTR1_ARATH^CTR1_ARATH^Q:2705-3418,H:562-805^49.59%ID^E:2e-69^RecName: Full=Serine/threonine-protein kinase CTR1 {ECO:0000303|PubMed:8431946}
YS2_GAGATTCC-ATAGAGGC_comp2275_seq1	NA	32	PPME1_MOUSE^PPME1_MOUSE^Q:263-1162,H:13-372^43.61%ID^E:2e-93^RecName: Full=Protein phosphatase methylesterase 1
YL1_CGCTCATT-TATAGCCT_comp13794_seq0	NA	33	.
YS1_GAGATTCC-TATAGCCT_comp3725_seq0	GBF1	33	CPRF1_PETCR^CPRF1_PETCR^Q:479-1702,H:1-406^54.33%ID^E:8e-136^RecName: Full=Common plant regulatory factor 1
YS1_GAGATTCC-TATAGCCT_comp19403_seq0	NA	33	.
YS1_GAGATTCC-TATAGCCT_comp15574_seq0	NA	33	CFI_CAMSI^CFI_CAMSI^Q:122-673,H:37-219^70.65%ID^E:2e-89^RecName: Full=Chalcone--flavonone isomerase
YL1_CGCTCATT-TATAGCCT_comp11427_seq0	NA	33	DCP1_ARATH^DCP1_ARATH^Q:257-907,H:1-211^77.52%ID^E:2e-108^RecName: Full=mRNA-decapping enzyme-like protein
AR2_GAGATTCC-GGCTCTGA_comp24852_seq0	NA	34	.
AR1_GAGATTCC-CCTATCCT_comp13415_seq0	NA	34	HFA6B_ARATH^HFA6B_ARATH^Q:459-1457,H:47-402^50.68%ID^E:5e-106^RecName: Full=Heat stress transcription factor A-6b
AR1_GAGATTCC-CCTATCCT_comp31298_seq0	NA	34	.
AR1_GAGATTCC-CCTATCCT_comp42375_seq0	NA	34	.
AR1_GAGATTCC-CCTATCCT_comp25166_seq0	NA	34	.
AR1_GAGATTCC-CCTATCCT_comp821_seq3	NA	34	PER5_VITVI^PER5_VITVI^Q:296-1213,H:26-328^55.56%ID^E:8e-105^RecName: Full=Peroxidase 5 {ECO:0000250|UniProtKB:P22195}
AR2_GAGATTCC-GGCTCTGA_comp15931_seq0	NA	34	PHT17_ORYSJ^PHT17_ORYSJ^Q:3-218,H:211-284^67.57%ID^E:9e-24^RecName: Full=Probable inorganic phosphate transporter 1-7
AR2_GAGATTCC-GGCTCTGA_comp370_seq0	NA	34	14KD_DAUCA^14KD_DAUCA^Q:321-569,H:54-136^67.47%ID^E:2e-28^RecName: Full=14 kDa proline-rich protein DC2.15
YL3_CGCTCATT-ATAGAGGC_comp101_seq3	redox2	34	MER_ERYCB^MER_ERYCB^Q:148-1101,H:4-327^54.77%ID^E:6e-114^RecName: Full=Methylecgonone reductase
AR1_GAGATTCC-CCTATCCT_comp1106_seq0	NA	34	CERK1_ARATH^CERK1_ARATH^Q:244-2034,H:22-617^59.74%ID^E:0^RecName: Full=Chitin elicitor receptor kinase 1
AR2_GAGATTCC-GGCTCTGA_comp12553_seq0	NA	34	LORF2_MOUSE^LORF2_MOUSE^Q:1716-3227,H:397-895^27.41%ID^E:2e-42^RecName: Full=LINE-1 retrotransposable element ORF2 protein
AR1_GAGATTCC-CCTATCCT_comp16225_seq0	NA	34	ACCH1_ARATH^ACCH1_ARATH^Q:118-960,H:12-288^50.18%ID^E:4e-73^RecName: Full=1-aminocyclopropane-1-carboxylate oxidase homolog 1

AR1_GAGATTCC-CCTATCCT_comp30027_seq0	NA	34	.
AR2_GAGATTCC-GGCTCTGA_comp22761_seq0	NA	34	.
YL3_CGCTCATT-ATAGAGGC_comp5165_seq0	NA	35	Y5102_ARATH^Y5102_ARATH^Q:1752-3065,H:1-410^75.8%ID^E:0^RecName: Full=Serine/threonine-protein kinase At5g01020
OL1_CGCTCATT-CCTATCCT_comp374_seq0	GS1	35	MTDH_FRAAN^MTDH_FRAAN^Q:212-1258,H:10-356^57.02%ID^E:1e-115^RecName: Full=Probable mannitol dehydrogenase
AR2_GAGATTCC-GGCTCTGA_comp10111_seq1	NA	35	PYRD2_RAT^PYRD2_RAT^Q:122-1753,H:34-581^51.72%ID^E:0^RecName: Full=Pyridine nucleotide-disulfide oxidoreductase domain-containing protein 2
AR1_GAGATTCC-CCTATCCT_comp696_seq16	NA	35	BGH3B_BACO1^BGH3B_BACO1^Q:1249-2835,H:107-658^31.05%ID^E:4e-72^RecName: Full=Beta-glucosidase BoGH3B
YS2_GAGATTCC-ATAGAGGC_comp6183_seq2	NA	35	YIF1B_MOUSE^YIF1B_MOUSE^Q:488-1138,H:83-305^32.61%ID^E:5e-28^RecName: Full=Protein YIF1B
YL1_CGCTCATT-TATAGCCT_comp10608_seq0	NA	35	SFT2B_HUMAN^SFT2B_HUMAN^Q:233-616,H:27-158^39.39%ID^E:4e-23^RecName: Full=Vesicle transport protein SFT2B
AR1_GAGATTCC-CCTATCCT_comp977_seq0	NA	35	SF3BA_ARATH^SF3BA_ARATH^Q:186-446,H:1-87^94.25%ID^E:8e-51^RecName: Full=Uncharacterized protein At3g23325
AR1_GAGATTCC-CCTATCCT_comp6_seq2	NA	36	.
AR1_GAGATTCC-CCTATCCT_comp24292_seq0	CrWRKY1	36	WRK70_ARATH^WRK70_ARATH^Q:551-766,H:109-180^62.5%ID^E:2e-21^RecName: Full=Probable WRKY transcription factor 70
AR1_GAGATTCC-CCTATCCT_comp3465_seq0	NA	36	Y3272_ARATH^Y3272_ARATH^Q:414-1277,H:85-371^69.79%ID^E:4e-151^RecName: Full=Kelch repeat-containing protein At3g27220
AR1_GAGATTCC-CCTATCCT_comp572_seq0	NA	36	.
AR2_GAGATTCC-GGCTCTGA_comp1660_seq1	BIS2	37	BH093_ARATH^BH093_ARATH^Q:717-1313,H:146-351^66.02%ID^E:2e-66^RecName: Full=Transcription factor bHLH93
YS2_GAGATTCC-ATAGAGGC_comp11524_seq0	NA	37	FBK9_ARATH^FBK9_ARATH^Q:114-1199,H:15-383^41.98%ID^E:4e-74^RecName: Full=F-box/kelch-repeat protein At1g23390
OL3_CGCTCATT-GGCTCTGA_comp23128_seq0	NA	37	.
OL3_CGCTCATT-GGCTCTGA_comp28247_seq0	NA	37	.
OL3_CGCTCATT-GGCTCTGA_comp55923_seq0	NA	37	.
OL3_CGCTCATT-GGCTCTGA_comp21836_seq0	NA	37	.
AR1_GAGATTCC-CCTATCCT_comp8890_seq1	NA	38	AGL8_SOLTU^AGL8_SOLTU^Q:160-879,H:1-250^68.9%ID^E:2e-113^RecName: Full=Agamous-like MADS-box protein AGL8 homolog
AR1_GAGATTCC-CCTATCCT_comp12530_seq0	D4H	38	DV4H_CATRO^DV4H_CATRO^Q:56-1183,H:24-400^67.72%ID^E:0^RecName: Full=Deacetoxyvindoline 4-hydroxylase
AR1_GAGATTCC-CCTATCCT_comp13728_seq0	NA	38	.
AR2_GAGATTCC-GGCTCTGA_comp26468_seq0	NA	39	.
AR1_GAGATTCC-CCTATCCT_comp18788_seq0	MATE1	39	TT12_ARATH^TT12_ARATH^Q:204-1712,H:3-492^47.34%ID^E:5e-141^.^.

Table S3

Primers	Sequence (5' - 3')	Application
vm16OMTlike-for	CTGAGAAGATCTATGGATGTGCAAAGTGATGAGTTTAG	**Heterologous expression of tabersonine 16-O-methyltransferase candidates in yeast**
vm16OMTlike-rev	CTGAGAAGATCTTTAAGGATAAACCTCAATGACACTTCTAATG	
yYS1_comp1836_seq0-for	CTGAGAGGATCCATGGATCTTGAAGCTTCTGAGATTCGT	
yYS1_comp1836_seq0-rev	CTGAGAGGATCCAGGATATACCTCAATAGGACACCTGAAATCTA	
vm16OMTYFP-for	CTGAGAACTAGTATGGATGTGCAAAGTGATGAGTTTAG	**Subcellular localization studies**
vm16OMTYFP-rev	CTGAGAACTAGTAGGATAAACCTCAATGACACTTCTAATG	
qVmActin-for	GGAGCTGAGAGATTCCGTTG	**qPCR analysis**
qVmActin-rev	GAATTCCTGCAGCTTCCATC	
qVm16OMT-for - (YS2_comp61_seq0)	CAGTTATCTATGGACATGGGCATGA	
qVm16OMT-rev - (YS2_comp61_seq0)	TTAGAAATTGAATTAGTACATTGTCCCT	
qVmYL1_comp255_seq0-for	CTACTACACTTCCTCTTAAGAGACGT	
qVmYL1_comp255_seq0-rev	ACTTTGGGGCAAAATATATTCCACTAA	
qVmOL3_comp1418_seq0-for	TGTAGAACTTGCATTGGCAAATTTGAT	
qVmOL3_comp1418_seq0-rev	TCTGTTCTCGTTTGTTGTTCCACTAT	
qVmOL3_comp1489_seq0-for	GATATGACAGAGAAATTTGGGTTGACA	
qVmOL3_comp1489_seq0-rev	ACGCTTTCGCCTGGAATAATAAGTG	
qVmAR1_comp3108_seq0-for	CAGAACACAATATAGTTGCAGACGTT	
qVAR1_comp3108_seq0-rev	ACATAGGAAACACTCTAAAATCTGCAT	
qVmYS2_comp528_seq2-for	TTCCATTTTGATTGGAAACTTTATGGT	
qVmYS2_comp528_seq2-rev	AGGCATGTGCAAAATGTTGCAATCC	

Part 3: Metabolic engineering of the vindoline biosynthetic pathway in yeast

As extensively presented on the course of this thesis, *Catharanthus roseus*, or Madagascar periwinkle is the source of high value molecules such as the anticancer vinblastine and vincristine, extensively used in chemotherapy treatments. The commercial production of these dimeric alkaloids relies mostly on the semisynthetic synthesis, where the natural monomeric alkaloids, catharanthine and vindoline, are condensed to form the vinblastine, and successively oxidized to form vincristine. Classically, catharanthine and vindoline are purified from *C. roseus* leaves, which considerably increases the price of the final product (around 50.000 dollars the gramme in the USA). In such a way, two papers were published regarding the production of the precursor vindoline in yeast cell factories in order to reach an industrial competitive level for vindoline production.

The first paper is focused on the resolution of two bottlenecks present on this pathway. The first being the production of the unwanted by-product vindorosine, issue of the enzymatic promiscuity of the pathway enzymes. The second, the methylation of 16-hydroxytabersonine which was found out as a problematic bottleneck, reducing the conversion of the subsequent reactions. In this publication, the technique of homologous recombination was used in order to create a stable strain that produces the vindoline intermediates.

The second paper presents the whole vindoline pathway optimized and inserted into the yeast via the CRISPR-Cas9 system. The final yeast strain was cultivated in a bioreactor and several parameters were optimized in order to achieve the best production rate of vindoline

Article 5 : Optimization of tabersonine methoxylation to increase vindoline precursor synthesis in yeast cell factories

In the article 5, I was responsible for the construction of the yeast strains and their bioconversion assays. The original mother strains (without the vindoline pathway genes) and the integration plasmid backbones were designed by other authors. The fusion proteins and the subcellular localisation studies were conducted by me, also all the metabolites quantification, statistical analysis and the writing of the manuscript, under the assistance of my superiors.

RESUME EN FRANÇAIS DE L'ARTICLE 5

Les métabolites spécialisés de plantes sont largement utilisés dans l'industrie pharmaceutique, notamment les alcaloïdes indoliques monoterpéniques (AIM) comme la vinblastine et la vincristine qui présentent toutes deux des propriétés anticancéreuses. Les deux composés peuvent être obtenus par condensation chimique de leurs précurseurs, la vindoline et la catharanthine extraites de feuilles de la pervenche de Madagascar. Cependant, l'utilisation intensive de ces molécules en chimiothérapie augmente la demande de précurseurs et entraîne des pénuries. C'est pourquoi le développement de productions alternatives utilisant des cellules-usines est requis. Dans ce contexte, la biosynthèse de vindoline à partir de tabersonine dans la levure exprimant les gènes hétérologues de plantes impliqués dans cette bioconversion présente un intérêt particulier bien qu'à ce jour, un tel système n'ait pas atteint des niveaux de production élevés. Afin de contourner les goulots d'étranglement dans la chaîne métabolique, le flux métabolique a été canalisé vers l'alcaloïde d'intérêt en modulant le nombre de copies des deux premiers gènes de la voie de biosynthèse de la vindoline, à savoir la tabersonine 16-hydroxylase et la tabersonine-16-O-méthyltransférase. L'augmentation du nombre de copies de gènes a entraîné une méthoxylation optimisée de la tabersonine. Elle a également contribué à résoudre le problème lié à la compétition pour l'accès à la tabersonine de la troisième enzyme de la voie, la tabersonine 3-oxygénase qui présente une promiscuité de substrat élevée. Grâce à cette approche, nous avons réussi à créer une souche de levure qui produit le quatrième intermédiaire biosynthétique de la voie de la vindoline sans accumuler d'autres intermédiaires ou de produits secondaires indésirables. Cette optimisation ouvre la voie au développement futur de levures utilisées comme cellules-usines pour produire la vindoline à une échelle industrielle.

molecules

MDPI

Article

Optimization of Tabersonine Methoxylation to Increase Vindoline Precursor Synthesis in Yeast Cell Factories

Pamela Lemos Cruz [1], Natalja Kulagina [1], Grégory Guirimand [1,2,3], Johan-Owen De Craene [1], Sébastien Besseau [1], Arnaud Lanoue [1], Audrey Oudin [1], Nathalie Giglioli-Guivarc'h [1], Nicolas Papon [4], Marc Clastre [1] and Vincent Courdavault [1,*]

1 EA2106 "Biomolécules et Biotechnologies Végétales", Université de Tours, 37000 Tours, France; pamela.lemos@univ-tours.fr (P.L.C.); natalja.kulagina@univ-tours.fr (N.K.); gregory.guirimand@univ-tours.fr (G.G.); johan-owen.decraene@univ-tours.fr (J.-O.D.C.); sebastien.besseau@univ-tours.fr (S.B.); arnaud.lanoue@univ-tours.fr (A.L.); audrey.oudin@univ-tours.fr (A.O.); nathalie.guivarch@univ-tours.fr (N.G.-G.); marc.clastre@univ-tours.fr (M.C.)
2 Graduate School of Sciences, Technology and Innovation, Kobe University, Kobe 657-8501, Japan
3 Le Studium Loire Valley Institute for Advanced Studies, 45000 Orléans & Tours, France
4 Univ Angers, Univ Brest, GEIHP, SFR ICAT, F-49000 Angers, France; nicolas.papon@univ-angers.fr
* Correspondence: vincent.courdavault@univ-tours.fr

check for updates

Citation: Lemos Cruz, P.; Kulagina, N.; Guirimand, G.; De Craene, J.-O.; Besseau, S.; Lanoue, A.; Oudin, A.; Giglioli-Guivarc'h, N.; Papon, N.; Clastre, M.; et al. Optimization of Tabersonine Methoxylation to Increase Vindoline Precursor Synthesis in Yeast Cell Factories. *Molecules* **2021**, *26*, 3596. https://doi.org/10.3390/molecules26123596

Academic Editors: Dezső Csupor, Fang-Rong Chang and Mohamed El-Shazly

Received: 15 April 2021
Accepted: 28 May 2021
Published: 11 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Plant specialized metabolites are widely used in the pharmaceutical industry, including the monoterpene indole alkaloids (MIAs) vinblastine and vincristine, which both display anticancer activity. Both compounds can be obtained through the chemical condensation of their precursors vindoline and catharanthine extracted from leaves of the Madagascar periwinkle. However, the extensive use of these molecules in chemotherapy increases precursor demand and results in recurrent shortages, explaining why the development of alternative production approaches, such microbial cell factories, is mandatory. In this context, the precursor-directed biosynthesis of vindoline from tabersonine in yeast-expressing heterologous biosynthetic genes is of particular interest but has not reached high production scales to date. To circumvent production bottlenecks, the metabolic flux was channeled towards the MIA of interest by modulating the copy number of the first two genes of the vindoline biosynthetic pathway, namely tabersonine 16-hydroxylase and tabersonine-16-*O*-methyltransferase. Increasing gene copies resulted in an optimized methoxylation of tabersonine and overcame the competition for tabersonine access with the third enzyme of the pathway, tabersonine 3-oxygenase, which exhibits a high substrate promiscuity. Through this approach, we successfully created a yeast strain that produces the fourth biosynthetic intermediate of vindoline without accumulation of other intermediates or undesired side-products. This optimization will probably pave the way towards the future development of yeast cell factories to produce vindoline at an industrial scale.

Keywords: alkaloids; *Catharanthus roseus*; metabolic engineering; O-methyltransferase

1. Introduction

Plant specialized metabolites were first described in the middle 1920s and were initially named secondary metabolites because these compounds were thought to be non-essential for plant development, in contrast to primary metabolites. Now, accumulating evidence has established that specialized metabolites allow plants to cope with their environment and notably contribute to the arsenal of defenses that plants have evolved against bioaggressors, making specialized metabolites essential components of plant life [1–3]. Biological activities of specialized metabolites also explain why they have been associated with medicinal plant properties commonly used in traditional medicine [4,5]. Several of these metabolites, such as the analgesic morphine, the antitussive codeine from *Papaver somniferum* [6], and the

antimalarial artemisinin from *Artemisia annua* [7], for instance, have been of great importance to modern medicine for decades. Along with other active plant-born compounds, they constitute around 25% of all the drugs currently used in human health [8].

Monoterpene indole alkaloids (MIAs) are one of the prominent examples of plant specialized metabolites used in our current pharmacopeia. Almost all MIAs derive from the common precursor strictosidine that is further metabolized to yield the plethora of MIAs accumulated in Apocynaceae, Rubiaceae, Nyssaceae, and Loganiaceae [9]. Among MIAs, vinblastine and vincristine, both specific to the Madagascar periwinkle (*Catharanthus roseus*) [10], are widely exploited in anticancer treatments [11–13]. The supply of these compounds mostly relies on plant culture and on the extraction of their two precursors, vindoline and catharanthine, from *C. roseus* leaves, which are further condensed to produce the active compounds (Figure 1A) [14,15]. This supply is obviously highly dependent on plant growth and suffers from recurrent shortages. On the other hand, the chemical synthesis of these molecules is laborious and inapplicable at an industrial scale due to their structural complexity [16–19], while in vitro cell cultures still fail to produce these MIAs [20].

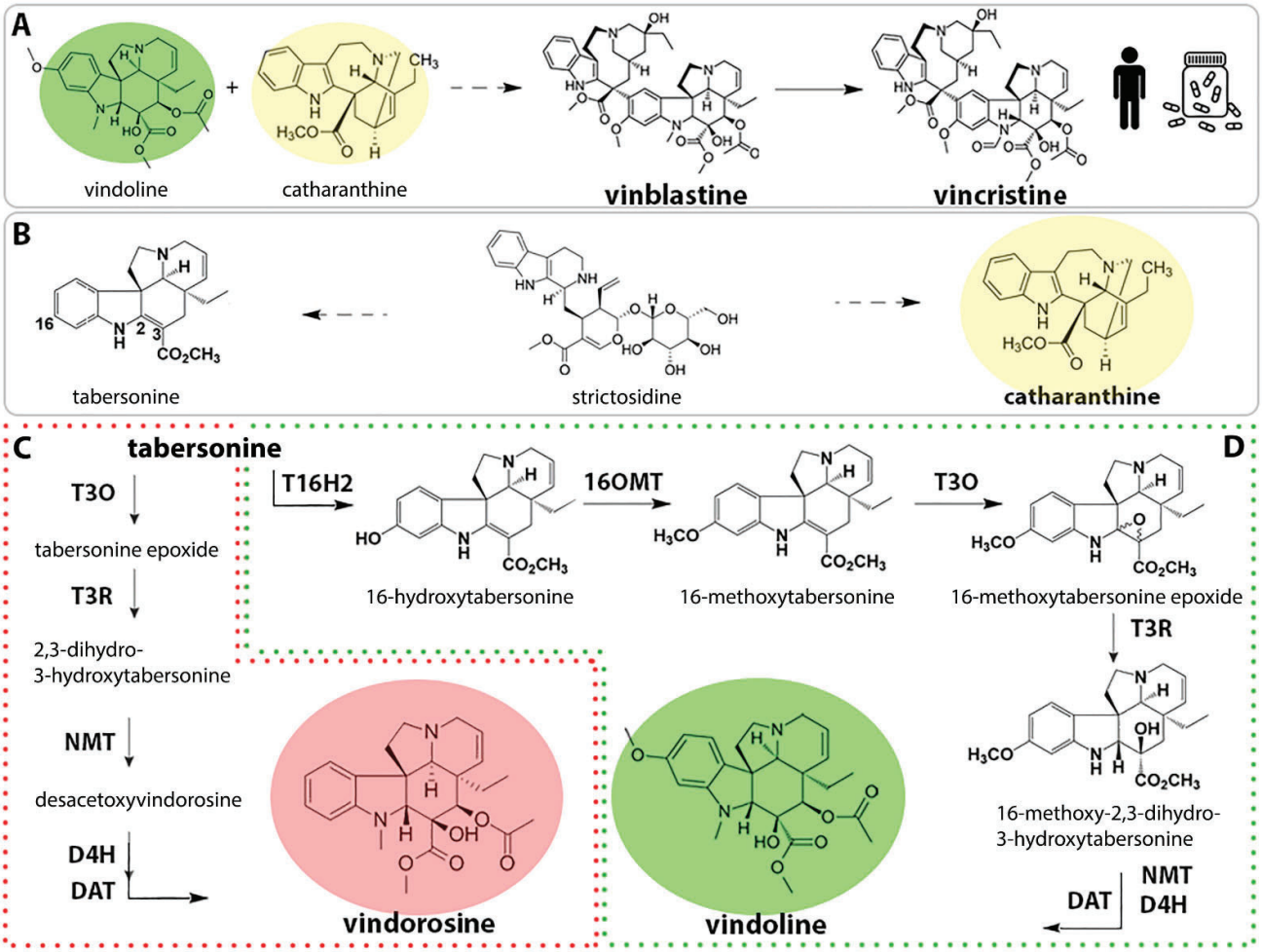


Figure 1. Vinblastine and vincristine production via semi-synthetic synthesis. (**A**) Illustration of the semi-synthetic synthesis of vinblastine and vincristine. (**B**) Tabersonine and catharanthine biosynthesis from strictosidine. (**C**,**D**) Tabersonine bioconversion in planta (**C**): by-product vindorosine biosynthesis. (**D**): vindoline biosynthesis). Solid line: one enzymatic step, discontinuous line: more than one enzymatic step.

Alternatively, the recent development of metabolic engineering strategies and heterologous productions provide new perspectives towards the supply of plant molecules of interest [21]. These production strategies basically rely on the reconstitution of a biosynthetic

pathway into a heterologous host through gene transfer. Among the potential heterologous hosts, yeast is considered as one of the most suitable organisms for metabolic engineering due to its fast growth, easy genetic manipulation, and available genome sequence [22]. Following the seminal heterologous productions of artemisinin [23], hydrocortisone [24], and progesterone [25], several plant alkaloids have been more recently biosynthesized by recombinant yeast, such as MIAs [26–28] but also benzylisoquinoline [29–32] and tetrahydroisoquinoline [33,34] alkaloids. However, heterologous biosynthesis of MIAs remains challenging due to the high complexity of the pathway and the elaborate cellular and subcellular compartmentalization of enzymes [35–37]. For instance, the central MIA precursor strictosidine was de novo produced in yeast at 0.5 mg/L [26], demonstrating the difficulty of reconstituting the entire metabolic pathway and obtaining high-scale production from glucose. By contrast, precursor-directed production, relying on yeast being fed highly abundant biosynthetic intermediates, represents an appealing alternative. Tabersonine is indeed an abundant MIA produced from strictosidine (Figure 1B) and accumulated in the seeds of *Voacanga africana* (25 to 30 g of tabersonine per kg of seed [38]). While tabersonine can be further metabolized into several derivatives, including, for instance, melodinine K [39], this compound is also converted into vindoline in *C. roseus* [40]. As such, tabersonine thus represents a highly valuable compound that can be used to deploy a precursor-directed synthesis of vindoline in engineered yeasts. However, although this bioconversion has been described in yeast [16], only a modest vindoline yield of 1.1 $mg \cdot L^{-1}$ 12 h^{-1} was reached, thus shedding light on the requirement of further optimizations of this strategy.

In *C. roseus* leaves, the tabersonine-to-vindoline conversion involves a biosynthetic route composed of seven steps [16]. Firstly, tabersonine is hydroxylated by tabersonine-16-hydroxylase (T16H2) to produce 16-hydroxytabersonine [41–43], followed by an *O*-methylation by tabersonine-16-*O*-methyltransferase (16OMT) [44,45]. The resulting 16-methoxytabersonine is then epoxidized by tabersonine 3-oxygenase (T3O) [46] and reduced by tabersonine 3-reductase (T3R) [16,45], generating the 16-methoxy-2,3-dihydro-3-hydroxytabersonine (Figure 1B). Finally, vindoline is produced through three further reactions of N-methylation [47], hydroxylation, and acetylation [48], catalyzed by 3-hydroxy-16-methoxy-2,3-dihydrotabersonine N-methyltransferase (NMT), desacetoxyvindoline-4-hydroxylase (D4H), and deacetylvindoline 4-*O*-acetyl transferase (DAT) [45,48,49], respectively (Figure 1C). Due to substrate promiscuity, T3O is also able to directly metabolize tabersonine to produce tabersonine epoxide, which is further converted into vindorosine and its biosynthetic intermediates by downstream enzymes T3R, NMT, D4H, and DAT [16,43,46] (Figure 1C).

Since vindorosine cannot be methoxylated a posteriori, vindorosine production affects, in turn, the synthesis of vinblastine and vincristine because it lacks the functional group involved in condensation with catharanthine [43]. Interestingly, a similar hijacking reaction was also observed in the engineered yeast expressing the vindoline pathway [16]. In these conditions, the production of vindorosine even exceeded vindoline synthesis and was accompanied by the massive accumulation of biosynthetic intermediates from both pathways. Therefore, the tight control of the metabolic flux in yeast constitutes a main concern for an optimal production of vindoline through tabersonine bioconversion with reduced accumulation of intermediates and limited vindorosine synthesis. A similar rationale has been already pointed out for benzylisoquinoline [29–32] and tetrahydroisoquinoline [33,34] alkaloids.

The present study aimed at optimizing the vindoline production through tabersonine conversion in yeast by resolving the bottleneck of early enzymatic steps to direct the metabolic flux towards vindoline. The first two enzymatic steps of the conversion were thus investigated by analyzing substrate permeability to yeast membranes, enzyme subcellular localization, enzyme isoform, and the effect of gene copy number to prevent tabersonine uptake by the competing T3O. By fine-tuning the copy number of T16H2 and 16OMT, we successfully orientated the conversion of tabersonine towards vindoline and limited

biosynthetic intermediate accumulation. In addition, we also drastically prevented the formation of vindorosine precursors, thus offering new perspectives towards the high production of vindoline in yeast cell factories and for other syntheses of natural products in yeast that suffer from endogenous substrate competition.

2. Results

2.1. Identification of Early Pathway Bottlenecks

To optimize the synthesis of vindoline in engineered yeast and reduce the biosynthetic flux towards the undesired vindorosine (Figure 1C), we first investigated the competition between T16H2 and T3O for tabersonine. Using galactose-inducible episomal vectors of expression, we compared the effect of the T16H2/T3O gene copy ratio on 16-hydroxytabersonine and tabersonine epoxide synthesis (Table 1).

Table 1. Yeast strains and the respective plasmids used for the transformation.

Strains	Plasmids
Inducible WAT11 strains	
T16H2+T3O	pYEDP60_T16H2, pESC-TRP_empty, pESC-HIS_empty, pESC-LEU_T3O
2(T16H2)+T3O	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_empty, pESC-LEU_T3O
2(T16H2)_T3O	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_empty, pESC-LEU_16OMT_T3O
16OMT_se *	pESC_LEU_16OMT
ER_16OMT	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_empty, pESC-LEU_EROMT
16OMT	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_16OMT, pESC-LEU_empty
2(16OMTs)	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_16OMT, pESC-LEU_16OMT
Vm16OMT	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_Vm16OMT, pESC-LEU_empty
2(Vm16OMT)	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_Vm16OMT, pESC-LEU_Vm16OMT
16OMT_Vm16OMT	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_16OMT, pESC-LEU_Vm16OMT
T16H	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS empty, pESC-LEU_empty
16OMT_T3O	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS empty, pESC-LEU_16OMT_T3O
2(16OMTs)_T3O	pYEDP60_T16H2, pESC-TRP_T16H2, pESC-HIS_16OMT, pESC-LEU_16OMT_T3O
Empty_vectors	pYEDP60_empty, pESC-TRP_empty, pESC-HIS_empty, pESC-LEU_empty
Stable JDC058_CPR strains	
stable_2(16OMT)s	pHISA_pGK1-T16H2-tPRM5_pTEF1-16OMT-tCPS1, pURAK_pTDH3-T16H2-tPRM5_pTEF1T3O-tIDP1,pLEUA_pTEF1-16OMT-tCPS1_pGK1-T3R-tHIS5

* 16OMT_se: 16OMT single expression.

One or two copies of T16H2 along with one copy of T3O were co-expressed in the *Saccharomyces cerevisiae* Wat11 strain fed with 250 µM of tabersonine. The MIA contents of the extracellular media were analyzed by UPLC-MS 24 h post-feeding (Supplemental Figure S1) since close to 95% of vindoline pathway intermediates are secreted into the culture medium [16]. In the presence of a single T16H2 copy, tabersonine was almost fully consumed (Figure 2). This led to the synthesis of both 16-hydroxytabersonine and tabersonine epoxide in a circa 70/30 ratio, in agreement with the higher affinity of T16H2 for tabersonine [33,36]. Interestingly, the addition of a second T16H2 copy resulted in a significantly increased consumption of tabersonine, accompanied by a 1.25-fold higher production of 16-hydroxytabersonine and a 4-fold decrease in tabersonine epoxide biosynthesis, making 16-hydroxytabersonine the most abundant compound by far (>90%; Figure 2). This demonstrated that increasing the T16H2 gene copy number versus T3O could constitute a promising strategy to reduce vindorosine formation.

Based on this first observation, we next co-expressed 16OMT from *C. roseus* using the same plasmid system in the yeast strain expressing two copies of T16H2 and one of T3O (Table 1). Twenty-four hours after tabersonine feeding, 16-methoxytabersonine epoxide, which results from the subsequent activity of T16H2/16OMT/T3O, appeared as the main MIA accumulated in the culture medium with low amounts of tabersonine epoxide (Figure 3). However, while low amounts of 16-methoxytabersonine were detected,

a high accumulation of 16-hydroxytabersonine was observed, reaching a similar order of magnitude as 16-methoxytabersonine epoxide. In addition, we noted that this metabolite accumulation remained stable over time, even 48 h after feeding. This suggested that methylation of 16-hydroxytabersonine could represent a concrete bottleneck that needs to be solved to improve vindoline production.

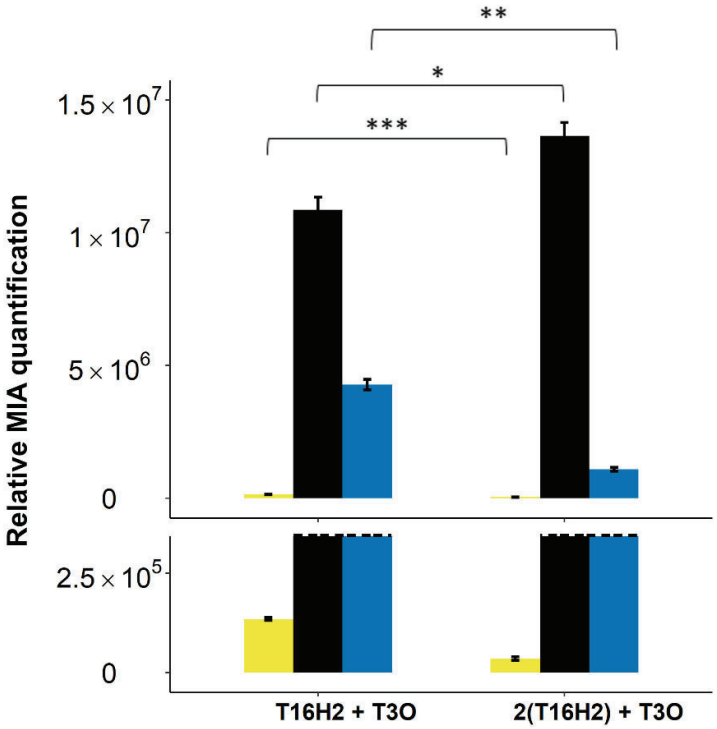


Figure 2. Evaluation of the competition between T16H2 and T3O for tabersonine in yeast. Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with tabersonine (250 µM). The dashed line represents the scale cut for the visualization of accumulated intermediates of low volume. The yeast strain harbors episomal plasmids containing one or two copies of T16H2 and one copy of T3O. Light yellow = tabersonine, black = 16-hydroxytabersonine, blue = tabersonine epoxide. Statistical analyses were performed with a two-tailed *t*-test ($p \leq 0.1$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$). MIA composition of the yeast culture medium is expressed as relative peak areas.

2.2. Evaluation of S. cerevisiae Cell Permeability to 16-Hydroxytabersonine

Since 16-hydroxytabersonine is highly accumulated in yeast culture medium, we next questioned whether such accumulation could result from a low permeability of this compound toward yeast membranes. As previously observed [16], and confirmed in this work, the vindoline biosynthetic intermediates are continuously excreted by yeast in the culture medium and re-internalized, allowing downstream MIA conversion. While T16H2 is anchored to the endoplasmic reticulum (ER, [43]), the methylation catalyzed by *C. roseus* 16OMT takes place in the cytosol and could be drastically reduced if 16-hydroxytabersonine remains in the culture medium ([37], Figure 4A). To test the capacity of 16-hydroxytabersonine import, yeasts were transformed with the *C. roseus* 16OMT-expressing plasmid or the corresponding empty vector (Table 1) and fed with 250 µM of 16-hydroxytabersonine for 24 h before analysis of the MIA content of the culture medium. While only 16-hydroxytabersonine was detected for the control strain transformed with the empty vector, the formation of 16-methoxytabersonine resulting from the methylation of 16-hydroxytabersonine was observed for the yeast strain expressing 16OMT (Figure 4B).

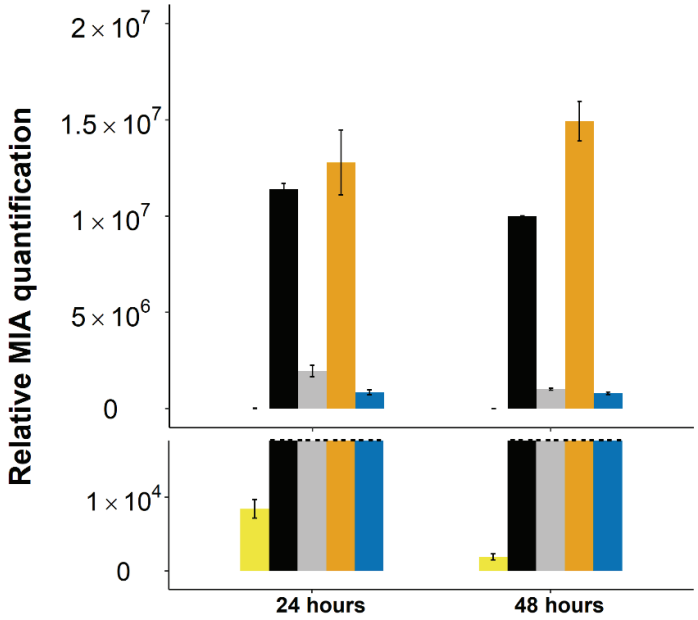


Figure 3. Evolution of MIA biosynthetic intermediates in the medium culture of yeast harboring episomal plasmids containing two copies of T16H2 and one copy of 16OMT and T3O (2(T16H2)_T3O strain). The dashed line represents the scale cut for the visualization of accumulated intermediates of low volume. Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with tabersonine (250 µM). Light yellow = tabersonine (tab), black = 16-hydroxytabersonine (16OHtab), grey = 16-methoxytabersonine (16MeOtab), dark yellow = 16-methoxytabersonine epoxide (16MeOtab_epoxi), blue = tabersonine epoxide. Error bars correspond to the standard error of biological replicates (n = 3). MIA composition of the yeast culture medium is expressed as relative peak areas.

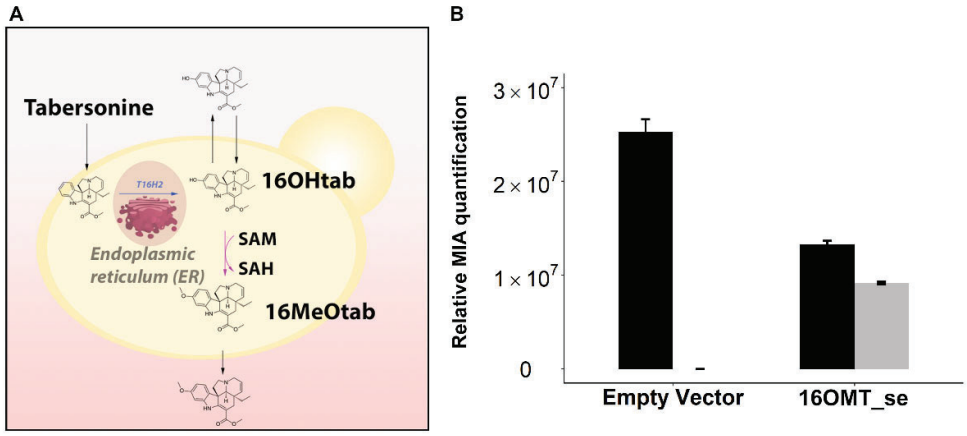


Figure 4. Cell permeability to 16-hydroxytabersonine (16OHtab). (**A**): Illustration of the pathway architecture in the inducible yeast strains from tabersonine to 16-methoxytabersonine (16MeOtab). (**B**): Permeability was evaluated by feeding yeasts (16OMT_se strain: episomal expression of 16OMT (se: single expression), Empty Vector: yeast strain harboring empty episomal plasmid) with 16-hydroxytabersonine. Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with 16-hydroxytabersonine (250 µM). Black = 16-hydroxytabersonine, grey = 16-methoxytabersonine. Error bars correspond to the standard error of biological replicates (n = 3). MIA composition of the yeast culture medium is expressed as relative peak areas.

This result thus showed that 16-hydroxytabersonine, originally fed outside of the cells, is accessible to 16OMT and established that *S. cerevisiae* cells are able to uptake 16-hydroxytabersonine from the culture medium. Consequently, this suggests that the

accumulation of 16-hydroxytabersonine is likely to be associated with an insufficient 16OMT activity inside yeast cells.

2.3. Optimization of 16OMT Activity in Yeast

2.3.1. Targeting 16OMT to the Endoplasmic Reticulum to Favor Metabolic Channeling

To promote 16OMT activity in yeast cells, we first investigated the effect of enzyme subcellular localization on the MIA biosynthetic flux. Targeting enzymes to specific organelles and enzyme physical associations have been tested previously to improve reaction rate and to avoid biosynthetic intermediate losses [50,51]. For instance, altering the localization of *S. cerevisiae* alcohol acyltransferase from the mitochondria to ER and intracellular lipid droplets enhanced 23-fold and 3-fold its expression and enzymatic activity, respectively [52]. Alternatively, insulating *Papaver somniferum* codeine reductase to the ER significantly increased enzyme specificity and morphine titer in yeast [53].

Given the distinct intracellular localizations of T16H2 and 16OMT in *C. roseus* cells [37], we tried to favor the metabolic channeling between both enzymes by prompting the anchoring of 16OMT to the ER. A chimeric 16OMT bearing the T16H2 transmembrane helix ([43]; Figure S2) was thus created and the resulting ER localization was confirmed in vivo (Figure S3). Coexpressing T16H2 in tandem with the ER-anchored 16OMT (ER_16OMT) was next achieved to potentially facilitate the metabolic channeling between both enzymes (Table 1; Figure 4A). As previously observed, an almost complete consumption of tabersonine was monitored for both strains (Figure 5). However, while 16 methoxytabersonine was produced at a level close to that of 16-hydroxytabersonine in the strain bearing a cytosolic 16OMT, a substantial decrease of the production of this compound was observed with the ER_16OMT. This lowering tended to suggest that ER anchoring altered 16OMT activity as a possible result of a decrease of 16OMT intrinsic activity or 16OMT enzyme amount. This strategy was thus not retained for methoxylation improvement.

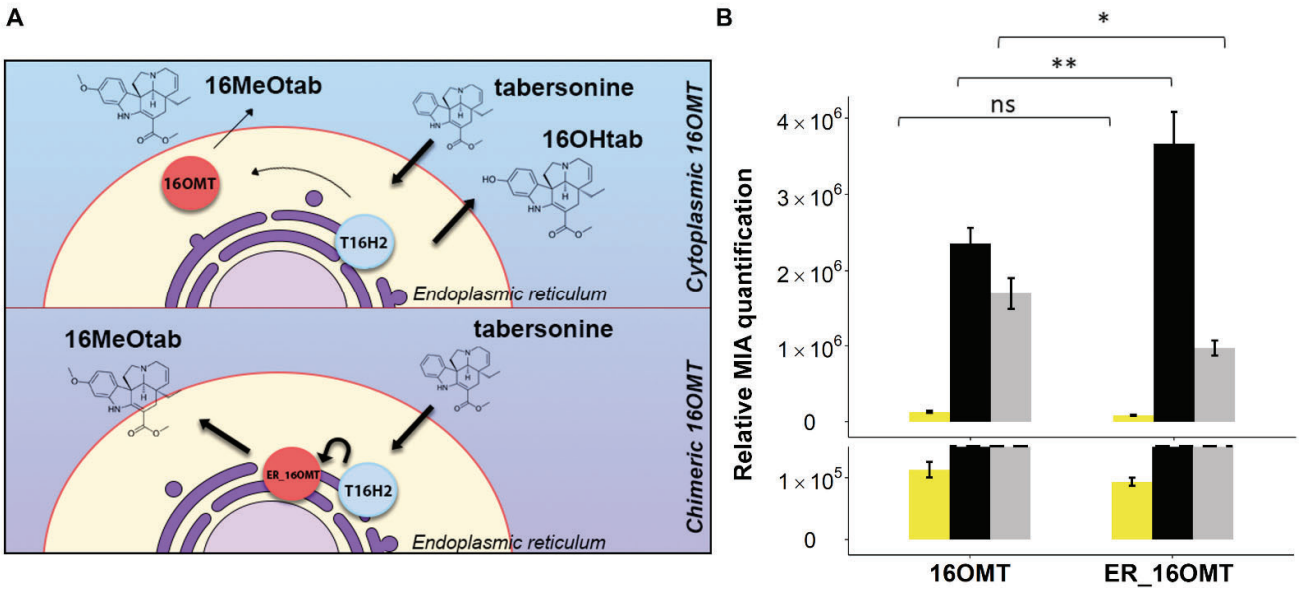


Figure 5. Localization approach for metabolic channeling. (**A**): Schematic illustration of ER-anchored 16OMT (ER_16OMT) construction and its hypothetical metabolic channeling. (**B**): The effect of anchoring 16OMT to ER was evaluated by feeding yeasts (16OMT strain: episomal expression of 2xT16H2 and native 16OMT; ER_16OMT strain: episomal expression of 2xT16H2 and ER_16OMT) with tabersonine. Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with tabersonine (250 µM). The dashed line represents the scale cut for the visualization of accumulated intermediates of low volume. Statistical analyses were performed with a two-tailed *t*-test ($p <= 0.1$, *: $p <= 0.05$, **: $p <= 0.01$, ns: not significant). Light yellow = tabersonine, black = 16-hydroxytabersonine, grey = 16-methoxytabersonine. Error bars correspond to the standard error of biological replicates ($n = 3$). MIA composition of the yeast culture medium is expressed as relative peak areas.

2.3.2. Testing a Distinct 16OMT Isoform and Increasing OMT Gene Copy Number to Limit 16-Hydroxytabersonine Accumulation

Promoting specialized metabolite synthesis in yeast can be achieved through the selection and expression of the most active orthologues of enzymes displaying low activity. A nice example of this strategy was recently described for phenylpyruvate reductase to optimize the synthesis of tropane alkaloids in heterologous hosts [31,32]. We thus took advantage of the recently published orthologue of 16OMT from *Vinca Minor* (Vm16OMT) and measured its 16-hydroxytabersonine methylation activity in yeast [54]. Vm16OMT was co-expressed using an inducible episomal plasmid along with two copies of T16H2 and compared to the similar strain bearing the *C. roseus* 16OMT, generated previously (Table 1). Again, tabersonine methoxylation was evaluated by measuring the amount of 16-methoxytabersonine in the culture medium. Surprisingly, while Vm16OMT and *C. roseus* 16OMT display similar catalytic properties [54], Vm16OMT seemed to be 4.5 times less active based on 16-methoxytabersonine accumulation (Figure 6). This decreased activity may have resulted from a low expression of Vm16OMT and suggested that this orthologue cannot substitute *C. roseus* 16OMT in our working conditions. While new orthologues can be tested after their identification, *C. roseus* 16OMT remains the best orthologue tested so far.

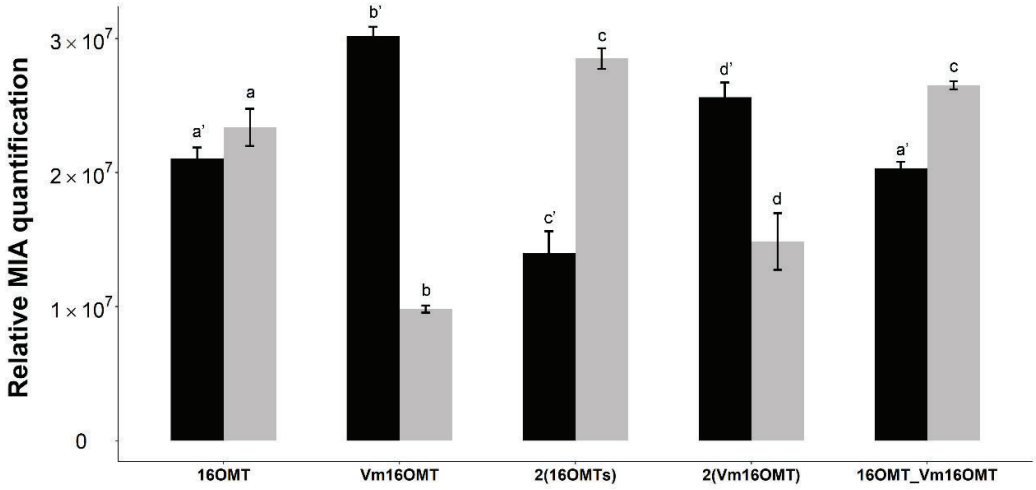


Figure 6. Comparison of 16OMT orthologue activities and effect of gene copy number on tabersonine methoxylation. Yeasts (16OMT strain: episomal expression of 2xT16H2 and 1x16OMT; Vm16OMT strain: episomal expression of 2xT16H2 and 1xVm16OMT; 2(16OMTs): episomal expression of 2xT16H2 and 2x16OMT; 2(Vm16OMT) strain: episomal expression of 2xT16H2 and 2xVm16OMT; 16OMT_Vm16OMT strain: episomal expression of 2xT16H2, 1x16OMT and 1xVm16OMT) were fed with tabersonine (250 µM) before analysis of the MIA content in the culture medium by UPLC-MS. Statistical analyses were performed with ANOVA followed by a Tukey test. Same letters express no significant difference between the means of the same MIA at 5% of significance: **a'**, **b'**, **c'**, **d'** = 16-hydroxytabersonine and **a**, **b**, **c**, **d** = 16-methoxytabersonine. Black = 16-hydroxytabersonine, grey = 16-methoxytabersonine. Error bars correspond to the standard error of biological replicates (n = 3). MIA composition of the yeast culture medium is expressed as relative peak areas.

Obviously, improving a low enzymatic activity in heterologous hosts can also be achieved by adjusting gene copy number as we did with T16H2 (Figure 2). As a consequence, an additional copy of *C. roseus* 16OMT was first expressed along with the two copies of T16H2. Interestingly, in these conditions, the accumulation of 16-hydroxytabersonine decreased down to circa 50% of the amount of the 16-methoxytabersonine accumulated in the culture medium compared to production with one 16OMT copy (Figure 6). Albeit less pronounced, a similar result was also obtained while expressing 2 copies of Vm16OMT versus a single Vm16OMT. Since *O*-methyltransferases operate as dimers, with heterodimers displaying modified kinetic parameters compared to homodimers, *C. roseus* 16OMT was also expressed in tandem with Vm16OMT. However, no improve-

ment of 16-methoxytabersonine production was observed in this condition compared to that of the yeast strain expressing two copies of *C. roseus* 16OMT. Overall, all these results established that increasing 16OMT gene copy number is an efficient strategy to limit 16-hydroxytabersonine accumulation, with *C. roseus* 16OMT being the most active orthologue.

Based on this observation, we next evaluated the impact of the expression of a second *C. roseus* 16OMT gene copy on the metabolic flux and the production of 16 methoxytabersonine epoxide. The yeast strain co-expressing two copies of both T16H2 and *C. roseus* 16OMT was further transformed to episomally express T3O (Table 1). The MIA content of the culture medium was analyzed 24 and 48 h after tabersonine feeding (Figure 7). In these conditions, the consumption of tabersonine was almost complete with a very low accumulation of tabersonine epoxide, confirming the positive effect of the two T16H2 copies. However, as compared to our original result (Figure 2), a different profile of downstream MIAs was observed at 24 h. We measured a main accumulation of 16-methoxytabersonine, confirming the increased capacity of the engineered yeasts to methylate 16-hydroxytabersonine. It did not result in an increased amount of 16-methoxytabersonine epoxide, thus suggesting that T3O activity could be limiting in this condition. In agreement with this hypothesis, an interesting evolution of the MIA content was monitored at 48 h (Figure 7). Firstly, we noted that 16-hydroxytabersonine was almost fully consumed and the resulting 16-methoxytabersonine followed a similar trend, indicating that both compounds continued to enter the biosynthetic flux. Secondly, we observed that 16-methoxytabersonine epoxide became by far the main accumulated MIA in contrast to our initial observation (Figure 2), suggesting that the conversion of tabersonine was of very high efficiency in these conditions.

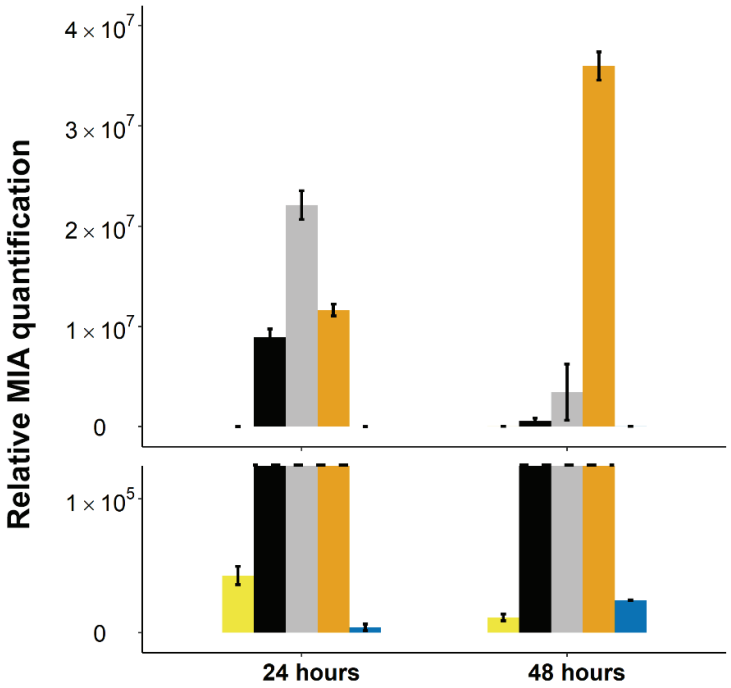


Figure 7. Evolution of MIA biosynthetic intermediates in the medium culture of yeast harboring episomal plasmids containing two copies of T16H2 and *C. roseus* 16OMT and one copy of T3O (2(16OMTs)_T3O strain). Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with tabersonine (250 µM). The dashed line represents the scale cut for the visualization of low accumulated intermediates. Error bars correspond to the standard error of biological replicates (n = 3). Light yellow = tabersonine, black = 16-hydroxytabersonine, grey = 16-methoxytabersonine, dark yellow = 16-methoxytabersonine epoxide, blue = tabersonine epoxide. MIA composition of the yeast culture medium is expressed as relative peak areas.

These results thus reinforced our previous finding regarding the positive effect of the second *C. roseus* 16OMT on tabersonine conversion. They also shed light on a possible

second limiting step catalyzed by T3O that necessitated more time to catalyze the reaction. This is in agreement with the poor T3O catalytic properties reported previously [16,46]. Above all, they confirmed that the optimization of tabersonine methoxylation has a positive effect on 16-methoxytabersonine epoxide biosynthesis.

2.4. Stable Integration of the Four First Genes of the Vindoline Pathway Paves the Way for High Vindoline Production

Finally, to improve the limiting activity of T3O compared to the high production of 16-methoxytabersonine, additional modifications of engineered yeast strains were performed. The relative proportion of several episomal plasmids can be highly variable and may affect enzyme populations and their respective activity. To prevent these negative effects, the first four genes of the vindoline pathway were thus stably integrated into the yeast genome. A set of integrative plasmids that allow recombination at classical metabolic markers was thus created along with a yeast strain invalidated for the corresponding metabolic genes, as described in the Methods section. We specifically chose the CEN.PK yeast strain since the potential of this strain is well recognized in metabolic engineering approaches [55]. Two copies of T16H2, *C. roseus* 16OMT and one copy of T3O and *C. roseus* cytochrome P450 reductase (CPR) were first integrated into the yeast genome (Table 1). Confirmation of gene integration in the resulting stable_2(16OMT)s strain was obtained by performing PCR on genomic DNA (Figure S4).

To limit the production of the undesired tabersonine epoxide, we decided not to increase the T3O gene copy number, which would have negatively imbalanced the competition between T16H2 and T3O for tabersonine access. Instead, we integrated a copy of T3R, the subsequent enzyme of the pathway, to consume the T3O product and potentially increase the biosynthetic flux. In this condition, 24 h after tabersonine feeding, we observed a huge consumption of tabersonine and a limited presence of 16-hydroxytabersonine, confirming that co-expressing two copies of T16H2 and *C. roseus* 16OMT prevents the accumulation of this compound (Figure 8). Only minute quantities of tabersonine epoxide and derivates (2,3-dihydro-3-hydroxytabersonine) were found. In addition, while 16-methoxytabersonine was still accumulated in high amounts, 16-methoxytabersonine epoxide was only detected at trace levels and 16-methoxy-2,3-dihydro-3-hydroxytabersonine became the most abundant MIA produced. Compared to our previous results (Figure 7), this suggested that addition of T3R may promote T3O activity, leading to an increased accumulation of downstream metabolites (Figure 8). This general trend was further confirmed at 48 h since the consumption of the biosynthetic intermediates continued to maximize the production of 16-methoxy-2,3-dihydro-3-hydroxytabersonine, which was 80% of the total produced MIAs (Figure 8; Figure S5).

Together, these results confirmed that the fine-tuning of gene copy number allows us to orient the metabolic flux toward the synthesis of the expected compound. For both yeast strains, stably or episomally expressing the vindoline pathway genes, altering the gene copy number reduced the accumulation of biosynthetic intermediates and avoided the hijacking of tabersonine for the synthesis of vindorosine precursors, thus addressing the first bottlenecks in vindoline precursor production.

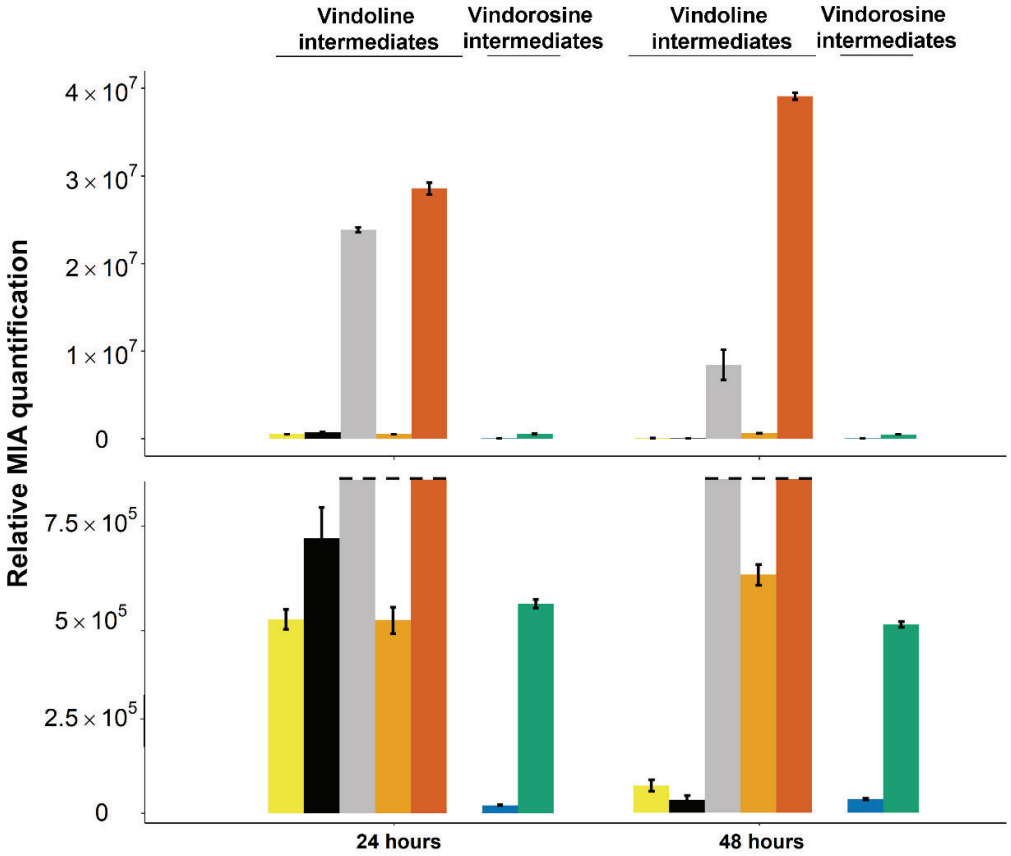


Figure 8. Evolution of MIA biosynthetic intermediates in the culture medium of yeast stably expressing two copies of T16H2 and *C. roseus* 16OMT and one copy of T3O and T3R (Stable_2(16OMT)s). Alkaloids were quantified by UPLC-MS in the yeast culture medium 24 h post-feeding with tabersonine (250 µM). The dashed line represents the scale cut for the visualization of low accumulated intermediates. Light yellow = tabersonine, black = 16-hydroxytabersonine, grey = 16-methoxytabersonine, dark yellow = 16-methoxytabersonine epoxide, orange = 16-methoxy-2,3-dihydro-3-hydroxytabersonine, blue = tabersonine epoxide, green = 2,3-dihydro-3-hydroxytabersonine. Error bars correspond to the standard error of biological replicates (n = 3). MIA composition of the yeast culture medium is expressed as relative peak areas.

3. Materials and Methods

3.1. Plasmid Construction

The galactose-inducible episomal vectors used in this study were pYeDP60 [56] and pESC vectors series purchased from Agilent (Santa Clara, CA, USA). All the genes cloned in pESC vectors were driven by *GAL10* promoter, except for T3O placed under *GAL1* promoter control (Table 1).

Integrative plasmids with bidirectional promoters were generated using pDONR221, pRS303, or pRS305 backbones. *S. cerevisiae* elements were PCR-amplified (Phusion™ High-Fidelity, ThermoFisher, Waltham, MA, USA) from *S. cerevisiae* gDNA. The promoters were amplified using specific primers containing overlap sequences (forward primers) to further create bidirectional pairs and *Spe*I/*Xba*I restriction sites (reverse primers) (Table S1) for downstream ORF cloning. The obtained DNA fragments were purified (PCR clean-up kit, Machery-Nagel, Düren, Germany) and combined by overlap PCR using promoter reverse primers. The plasmid pURAK (pDONR221 backbone) was constructed by cloning the bidirectional promoter pair of *S. cerevisiae* glycolytic genes *TEF1/TDH3* between *Spe*I and *Xba*I sites, and terminators of the *IDP1* gene between *Sac*I and *Spe*I, and the *PRM5* gene between *Xba*I and *Xho*I. The *URA3* gene was cloned in the *Pvu*II site. The plasmid pHISA (pRS303 backbone) was generated by cloning the bidirectional promoter pair of glycolytic genes *TEF1/PGK1* between *Spe*I and *Xba*I sites, and terminators of the *CPS1* gene between *Sac*I and *Spe*I and the *PRM5* gene between *Xba*I and *Xho*I. The plasmid pLEUA (pRS305

backbone) was constructed by cloning the bidirectional promoter pair of glycolytic genes *TEF1/PGK1* between *Spe*I and *Xba*I sites and terminators of the *CPS1* gene between *Sac*I and *Spe*I and the *HIS5* gene between *Xba*I and *Xho*I.

The plasmid pJDC1144 was created by cloning the *ARG3* gene in the *Eco*RV site of pDONR221, making a *Nco*I-*Eco*RV deletion in the *ARG3* and finally cloning the *URA3* gene in the *Pvu*II site. The plasmid pJDC256 was created by cloning the *TDH3* gene promoter and the *CYC1* gene terminator in the *Pvu*II site of plasmid pDONR221 and cloning the *C. roseus CPR2* gene optimized for yeast expression by Eurofins (Luxembourg, Luxembourg) in the *Bam*HI-*Eco*RI sites.

The genes of interest were amplified by PCR (Phusion™ High-Fidelity, ThermoFisher) using the primers containing *Spe1*, *BamHI*, *XbaI*, or *Nhe1* restriction sites for downstream ORF cloning (Table S1), followed by restriction enzyme digestion and ligation into the selected plasmids. The cDNA from leaves of *C. roseus* variant 'Little Bright Eye' was used as a template. The cDNA was produced from total *C. roseus* RNA following the RevertAid Reverse Transcriptase manufacturer's instructions (ThermoFisher).

Vector construction and ORF cloning were performed following standard molecular biology methods, which were carried out with *Escherichia coli* TOP10 cells and Luria-Bertani culture medium supplemented with the respective antibiotic for transformant selection (100 μg/mL ampicillin or 50 μg/mL kanamycin).

3.2. Yeast Strains

For the galactose-inducible vector system, *S. cerevisiae* WAT11 strain was employed, where the cytochrome P450 reductase gene from *Arabidopsis thaliana* (AtR1) was integrated into the genome [56].

For the integrative vector system, CEN.PK2-1C (MAT*a ura3-52 his3Δ1 leu2-3/112 trp1-289 MAL2-8^c SUC2*) strain (Euroscarf, Oberursel, Germany) was transformed by *Bgl*II-linearized pJDC144 to generate the *arg3* mutant, creating the strain JDC058 after excision of the plasmid by selecting yeast clones auxotrophic for uracil and arginine. This strain was further transformed by pJDC256, which carries the yeast optimized *C. roseus CPR* open reading frame (ORF) under the control of the *TDH3* gene promoter linearized by *Nco*I for the insertion at the *ARG3* locus (JDC058_CPR strain). The strains generated downstream are listed in Table 1.

3.3. Yeast Transformations and Culture

For the generation of *S. cerevisiae* WAT11 inducible strains, yeast competent cells were prepared prior to transformation [57]. The competent cells were transformed by electroporation with expression plasmids according to Table 1. The yeast selection was done by synthetic complete drop-out (SC) plates containing 0.67% yeast nitrogen base, 2% agar, 2% dextrose, and 0.05% DOB according to vector markers. Yeast overnight pre-cultures were performed in drop-out selection liquid medium for 16 h followed by the induction in YPGal medium (1% bactopeptone, 1% yeast extract, and 2% Gal) for 5 h at 28 °C with constant agitation of 200 rpm prior to the feeding with 250 μM of tabersonine (ChromaDex, Los Angeles, CA, USA) or 16-hydroxytabersonine as described by [43].

To generate *S. cerevisiae* CEN.PK stable strains, the previously constructed JDC058-CPR strain was transformed following the LiAc/PEG method [58] with the linearized (*StuI* for pURAK, *NheI* for pHISA, and *EcoRV* for pLEUA) integrative plasmids (Table 1). Selection of transformants was performed on synthetic complete drop-out (SC) plates according to the auxotrophic markers. Stable strains were further inoculated in liquid yeast extract-peptone-dextrose medium (YPD, 10 g L^{-1} of yeast extract, 20 g L^{-1} of peptone and 20 g L^{-1} of glucose) for 16 h under constant agitation of 200 rpm, followed by the 20-fold dilution in fresh YPD medium and feeding with 250 μM of tabersonine (ChromaDex) or 16-hydroxytabersonine [43] in the final volume of 200 μL.

Gene integration in the stable_2(16OMT)s yeast strain was confirmed by PCR amplifications performed on yeast genomic DNA extracted with the kit Plant II (Macherey-Nagel)

following the manufacturer's instructions. The wild type CEN.PK genomic DNA was also extracted to be used as a control. PCR reactions (Phusion™ High-Fidelity, ThermoFisher) were conducted using both DNAs as template and primers, as described in Table S1, before analysis in 1% agarose gels.

3.4. Construction of the Chimeric 16OMT

Prior to the construction of chimeric 16OMT, the transmembrane helix was amplified by PCR (Phusion™ High-Fidelity, ThermoFisher) using T16H2helix_FOR and T16H2helix_REV primers (Table S1) and pYEDP60_T16H2 plasmid as a template. The obtained fragment was further fused with the amplification product of 16OMT by PCR using the T16H2helix_FOR and the internal primer helix_OMT_REV (Table S1).

3.5. Yeast Metabolite Analysis

The bioconversion assays were followed by the separation of yeast extracellular media via centrifugation (8000 *g* for 10 min) and 18-fold dilution of supernatant in 100% methanol. The diluted samples were analyzed by UPLC-MS after 15 min of centrifugation at 16,000× *g*. The alkaloids were identified in the selected ion-monitoring mode as described in [59] (Figure S1). The results were processed using the QuanLynx™ software (Waters, Milford, MA, USA) in peak areas (ions count).

3.6. Subcellular Localization in C. roseus Cells

The subcellular localization of the chimeric 16OMT was confirmed by cloning the fusion PCR product into pSC-A plasmids. The cloning was performed in *SpeI* restriction sites in order to position YFP at the C-terminus of the chimeric protein while the ER-helix peptide signal was located at the N-terminus. The generated vector was further used for transient transformation of *C. roseus* cells by particle bombardment, and the images were prepared according to [37]. The vectors were co-transformed with plasmids expressing the CFP-ER (CD3-954) from the ABRC (http://www.arabidopsis.org, accessed on 10 June 2021) or CFP-nucleocytoplasm obtained from [37].

4. Conclusions

While the supply of vinblastine and vincristine suffers from recurrent shortages [60], there is an urgent need to develop alternative production approaches to maintain a continuous access to these essential drugs. Using yeast cell factories to produce the precursors of these compounds is one of the promising alternatives. A first proof-of-concept of vindoline production in yeast was described through precursor directed synthesis [16]. However, the low vindoline biosynthesis and the hijacking of the precursor tabersonine towards the production of the unwanted vindorosine indicated that further optimizations were required. In this context, we have shown here that tabersonine methoxylation constitutes a real bottleneck in vindoline precursor synthesis due to (i) the competition between tabersonine hydroxylation and epoxidation and (ii) the accumulation of 16-hydroxtabersonine over time. By adjusting the copy number of the first two genes of the vindoline biosynthetic pathway, namely T16H2 and *C. roseus* 16OMT, we successfully addressed these two problems and created a yeast strain producing the fourth biosynthetic intermediate of vindoline without massive accumulation of other intermediates or undesired side-products. This optimization of tabersonine methoxylation will probably pave the way towards the future development of yeast cell factories producing vindoline at an industrial scale. In addition, it could also guide the heterologous reconstitution of other biosynthetic pathways in yeast involving two successive steps of hydroxylation and methylation. However, after this first optimization, the balance of gene copy will probably need to be readjusted in the context of the full vindoline pathway reconstitution since new bottlenecks resulting from increased synthesis of early precursors may appears later in the pathway. Future studies will have to address this concerted gene expression at the whole pathway level and improve the intrinsic activity of the last enzymes of the pathway. Such an overall increase

of enzyme activity will probably involve the expression of additional proteins involved in the regeneration of some MIA pathway enzyme cofactors such as NADPH [22]. Delocalization of MIA biosynthetic enzymes to new subcellular compartments such as vacuoles or peroxisomes may also represent an interesting option to maximize the metabolic flux or avoid the accumulation of undesired/toxic intermediates [31,51,61]. Finally, controlled fed-batch fermentations of the newly developed yeast strains in bioreactors will probably increase MIA synthesis up to industrial scales.

Supplementary Materials: The following are available online, Table S1: List of primers, Figure S1: UPLC-MS/MS chromatograms of the natural products produced by the yeast strains. Figure S2: Fusion of the T16H2 transmembrane helix to the N-terminal end of 16OMT (ER_16OMT). Alignment of the first 55 residues of T16H2 with ER_16OMT. The red rectangle highlights the added sequence including the predicted transmembrane helix of identified in T16H2. Figure S3: Subcellular localization of 16OMT and EROMT in C. roseus cells. Cells were transformed transiently with 16OMT-YFP (A–D) and EROMT-YFP (E-H) expressing vectors in combination with CFP-nucleocytosolic or CFP-ER marker (second column). Co-localization of the two fluorescence signals appeared in the merged image (C, G). The morphology is observed with differential interference contrast (DIC). Bar: 10 um. Figure S4: Phusion PCR amplification of the integrated genes from the vindoline's pathway using genomic DNA from the designed stable yeast (Stable_2(16OMT)s) and genomic DNA from wild type CEN.PK (CEN.PK WT). The yeast gene SAM2 (S-adenosylmethionine synthetase) was used as positive control. T16H2: tabersonyne-16-hydroxylase, 16OMT: tabersonine-16-O-methyltransferase, T3O: tabersonine 3-oxygenase, T3R: tabersonine 3-reductase, CPR: optimized C. roseus CPR, SAM2: S-adenosylmethionine synthetase. Figure S5: Evolution of the accumulation of vindoline and vindorosine biosynthetic intermediates in the stable_2(16OMT)s yeast strain fed with tabersonine. Alkaloids were quantified by UPLC-MS in the yeast culture medium before and 24 and 48 hours post-feeding with tabersonine (250 μM). Error bars correspond to the standard error of biological replicates (n = 3).

Author Contributions: Investigation, P.L.C., N.K., G.G., J.-O.D.C., S.B., A.L., A.O., N.P.; data curation, N.G.-G., M.C.; writing—original draft preparation, P.L.C., N.K., V.C.; writing—review and editing, P.L.C., N.K., N.P., V.C.; supervision, M.C., V.C.; project administration, V.C.; funding acquisition, V.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Région Centre-Val de Loire, BioPROPHARM, CatharSIS and ETOPOCentre grants.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article.

Acknowledgments: G.G. and V.C. acknowledge the Studium Institute for the Studium Research Fellowship attributed to G.G.

Conflicts of Interest: The authors declare no conflict of interest.

Sample Availability: Samples of the compounds (16-hydroxytabersonine, 16-methoxytabersonine) are available from the authors.

References

1. Takshak, S.; Agrawal, S. Defense potential of secondary metabolites in medicinal plants under UV-B stress. *J. Photochem. Photobiol. B Biol.* **2019**, *193*, 51–88. [CrossRef] [PubMed]
2. Li, Y.; Kong, D.; Fu, Y.; Sussman, M.R.; Wu, H. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiol. Biochem.* **2020**, *148*, 80–89. [CrossRef]
3. Erb, M.; Kliebenstein, D.J. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* **2020**, *184*, 39–52. [CrossRef]
4. Favarin, D.C.; De Oliveira, J.R.; De Oliveira, C.J.F.; Rogerio, A.D.P. Potential Effects of Medicinal Plants and Secondary Metabolites on Acute Lung Injury. *BioMed Res. Int.* **2013**, *2013*, 576479. [CrossRef] [PubMed]
5. Wink, M. Modes of Action of Herbal Medicines and Plant Secondary Metabolites. *Medicines* **2015**, *2*, 251–286. [CrossRef]
6. Bernath, J. *The Genus Papaver*; Harwood Academic Publishers: Amsterdam, The Netherlands; CRC Press: Boca Raton, FL, USA, 2000; pp. 169–170, ISBN 9789057022715.

7. Paddon, C.J.; Keasling, J. Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Genet.* **2014**, *12*, 355–367. [CrossRef] [PubMed]
8. Yun, B.-W.; Yan, Z.; Amir, R.; Hong, S.; Jin, Y.-W.; Lee, E.-K.; Loake, G. Plant natural products: History, limitations and the potential of cambial meristematic cells. *Biotechnol. Genet. Eng. Rev.* **2012**, *28*, 47–60. [CrossRef] [PubMed]
9. St-Pierre, B.; Besseau, S.; Clastre, M.; Courdavault, V.; Courtois, M.; Crèche, J.; Ducos, E.; De Bernonville, T.D.; Dutilleul, C.; Glevarec, G.; et al. Deciphering the Evolution, Cell Biology and Regulation of Monoterpene Indole Alkaloids. In *Advances in Botanical Research*; Elsevier BV: Amsterdam, The Netherlands, 2013; Volume 68, pp. 73–109.
10. Srivastava, V.; Negi, A.S.; Kumar, J.; Gupta, M.; Khanuja, S.P. Plant-based anticancer molecules: A chemical and biological profile of some important leads. *Bioorg. Med. Chem.* **2005**, *13*, 5892–5908. [CrossRef]
11. Alam, M.M.; Naeem, M.; Khan, M.M.A.; Uddin, M. Vincristine and vinblastine anticancer catharanthus alkaloids: Pharmacological applications and strategies for yield improvement. In *Catharanthus Roseus*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 277–307.
12. Noble, R.L. The discovery of the vinca alkaloids—chemotherapeutic agents against cancer. *Biochem. Cell Biol.* **1990**, *68*, 1344–1351. [CrossRef]
13. Thomas, D.A.; Sarris, A.H.; Cortes, J.; Faderl, S.; O'Brien, S.; Giles, F.J.; Garcia-Manero, G.; Rodríguez, M.A.; Cabanillas, F.; Kantarjian, H. Phase II study of sphingosomal vincristine in patients with recurrent or refractory adult acute lymphocytic leukemia. *Cancer* **2005**, *106*, 120–127. [CrossRef] [PubMed]
14. Yu, F.; De Luca, V. ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in Catharanthus roseus. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 15830–15835. [CrossRef] [PubMed]
15. De Luca, V.; Laflamme, P. The expanding universe of alkaloid biosynthesis. *Curr. Opin. Plant Biol.* **2001**, *4*, 225–233. [CrossRef]
16. Qu, Y.; Easson, M.; Froese, J.; Simionescu, R.; Hudlicky, T.; De Luca, V. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6224–6229. [CrossRef] [PubMed]
17. Gueritte, F.; Bac, N.V.; Langlois, Y.; Potier, P. Biosynthesis of antitumour alkaloids from Catharanthus roseus. Conversion of 20′-deoxyleurosidine into vinblastine. *J. Chem. Soc. Chem. Commun.* **1980**, *10*, 452–453. [CrossRef]
18. O'Connor, S.E. Strategies for Engineering Plant Natural Products. *Biomembr. Part K Membr. Biog. Assem. Target. (Prokaryotes Mitochondria Chloroplasts)* **2012**, *515*, 189–206. [CrossRef]
19. Zhou, Y.; Gao, W.; Rong, Q.; Jin, G.; Chu, H.; Liu, W.; Yang, W.; Zhu, Z.; Li, G.; Zhu, G.; et al. Modular Pathway Engineering of Diterpenoid Synthases and the Mevalonic Acid Pathway for Miltiradiene Production. *J. Am. Chem. Soc.* **2012**, *134*, 3234–3241. [CrossRef] [PubMed]
20. Datta, A.; Srivastava, P. Variation in vinblastine production by Catharanthus roseus, during in vivo and in vitro differentiation. *Phytochemistry* **1997**, *46*, 135–137. [CrossRef]
21. Courdavault, V.; O'Connor, S.E.; Oudin, A.; Besseau, S.; Papon, N. Towards the Microbial Production of Plant-Derived Anticancer Drugs. *Trends Cancer* **2020**, *6*, 444–448. [CrossRef] [PubMed]
22. Guirimand, G.; Kulagina, N.; Papon, N.; Hasunuma, T.; Courdavault, V. Innovative Tools and Strategies for Optimizing Yeast Cell Factories. *Trends Biotechnol.* **2021**, *39*, 488–504. [CrossRef]
23. Paddon, C.J.; Westfall, P.J.; Pitera, D.J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M.D.; Tai, A.; Main, A.; Eng, D.; et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nat. Cell Biol.* **2013**, *496*, 528–532. [CrossRef] [PubMed]
24. Szczebara, F.M.; Chandelier, C.; Villeret, C.; Masurel, A.; Bourot, S.; Duport, C.; Blanchard, S.; Groisillier, A.; Testet, E.; Costaglioli, P.; et al. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat. Biotechnol.* **2003**, *21*, 143–149. [CrossRef] [PubMed]
25. Duport, C.; Spagnoli, R.; Degryse, E.; Pompon, D. Self-sufficient biosynthesis of pregnenolone and progesterone in engineered yeast. *Nat. Biotechnol.* **1998**, *16*, 186–189. [CrossRef] [PubMed]
26. Brown, S.; Clastre, M.; Courdavault, V.; O'Connor, S.E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 3205–3210. [CrossRef] [PubMed]
27. Carqueijeiro, I.; De Bernonville, T.D.; LaNoue, A.; Dang, T.; Teijaro, C.; Paetz, C.; Billet, K.; Mosquera, A.; Oudin, A.; Besseau, S.; et al. A BAHD acyltransferase catalyzing 19- O -acetylation of tabersonine derivatives in roots of Catharanthus roseus enables combinatorial synthesis of monoterpene indole alkaloids. *Plant J.* **2018**, *94*, 469–484. [CrossRef] [PubMed]
28. Carqueijeiro, I.T.; Brown, S.; Chung, K.; Dang, T.-T.T.; Walia, M.; Besseau, S.; De Bernonville, T.D.; Oudin, A.; LaNoue, A.; Billet, K.; et al. Two Tabersonine 6,7-Epoxidases Initiate Lochnericine-Derived Alkaloid Biosynthesis in Catharanthus roseus. *Plant Physiol.* **2018**, *177*, 1473–1486. [CrossRef]
29. Hawkins, K.M.; Smolke, C.D. Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat. Chem. Biol.* **2008**, *4*, 564–573. [CrossRef] [PubMed]
30. Li, Y.; Smolke, C.D. Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nat. Commun.* **2016**, *7*, 12137. [CrossRef] [PubMed]
31. Srinivasan, P.; Smolke, C.D. Biosynthesis of medicinal tropane alkaloids in yeast. *Nat. Cell Biol.* **2020**, *585*, 1–6. [CrossRef]
32. Courdavault, V.; Cassereau, J.; Papon, N. Engineered Microbes for Producing Anticholinergics. *ChemBioChem* **2021**, *22*, 1368–1370. [CrossRef]

33. Pyne, M.E.; Kevvai, K.; Grewal, P.S.; Narcross, L.; Choi, B.; Bourgeois, L.; Dueber, J.E.; Martin, V.J.J. A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nat. Commun.* **2020**, *11*, 1–10. [CrossRef] [PubMed]
34. Kulagina, N.; Papon, N.; Courdavault, V. Microbial Cell Factories for Tetrahydroisoquinoline Alkaloid Production. *ChemBioChem* **2021**, *22*, 639–641. [CrossRef] [PubMed]
35. Courdavault, V.; Papon, N.; Clastre, M.; Giglioli-Guivarc'H, N.; St-Pierre, B.; Burlat, V. A look inside an alkaloid multisite plant: The Catharanthus logistics. *Curr. Opin. Plant Biol.* **2014**, *19*, 43–50. [CrossRef] [PubMed]
36. De Bernonville, T.D.; Clastre, M.; Besseau, S.; Oudin, A.; Burlat, V.; Glévarec, G.; Lanoue, A.; Papon, N.; Giglioli-Guivarc'H, N.; St-Pierre, B.; et al. Phytochemical genomics of the Madagascar periwinkle: Unravelling the last twists of the alkaloid engine. *Phytochemistry* **2015**, *113*, 9–23. [CrossRef] [PubMed]
37. Guirimand, G.; Guihur, A.; Poutrain, P.; Héricourt, F.; Mahroug, S.; St-Pierre, B.; Burlat, V.; Courdavault, V. Spatial organization of the vindoline biosynthetic pathway in Catharanthus roseus. *J. Plant Physiol.* **2011**, *168*, 549–557. [CrossRef] [PubMed]
38. Poisson, J.E. Process of Obtaining Tabersonine. U.S. Patent 3758478A, 16 June 1973.
39. Walia, M.; Teijaro, C.N.; Gardner, A.; Tran, T.; Kang, J.; Zhao, S.; O'Connor, S.E.; Courdavault, V.; Andrade, R.B. Synthesis of (−)-Melodinine K: A Case Study of Efficiency in Natural Product Synthesis. *J. Nat. Prod.* **2020**, *83*, 2425–2433. [CrossRef]
40. O'Connor, S.E.; Maresh, J.J. Chemistry and biology of monoterpene indole alkaloid biosynthesis. *Nat. Prod. Rep.* **2006**, *23*, 532–547. [CrossRef]
41. St-Pierre, B.; De Luca, V. A Cytochrome P-450 Monooxygenase Catalyzes the First Step in the Conversion of Tabersonine to Vindoline in Catharanthus roseus. *Plant Physiol.* **1995**, *109*, 131–139. [CrossRef]
42. Schröder, G.; Unterbusch, E.; Kaltenbach, M.; Schmidt, J.; Strack, D.; De Luca, V.; Schröder, J. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: Tabersonine 16-hydroxylase. *FEBS Lett.* **1999**, *458*, 97–102. [CrossRef]
43. Besseau, S.; Kellner, F.; Lanoue, A.; Thamm, A.M.; Salim, V.; Schneider, B.; Geu-Flores, F.; Höfer, R.; Guirimand, G.; Guihur, A.; et al. A Pair of Tabersonine 16-Hydroxylases Initiates the Synthesis of Vindoline in an Organ-Dependent Manner in Catharanthus roseus. *Plant Physiol.* **2013**, *163*, 1792–1803. [CrossRef]
44. Levac, D.; Murata, J.; Kim, W.S.; De Luca, V. Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of Catharanthus roseus 16-hydroxytabersonine-16-O-methyltransferase. *Plant J.* **2007**, *53*, 225–236. [CrossRef]
45. De Luca, V.; Fernandez, J.A.; Campbell, D.; Kurz, W.G.W. Developmental Regulation of Enzymes of Indole Alkaloid Biosynthesis in Catharanthus roseus. *Plant Physiol.* **1988**, *86*, 447–450. [CrossRef] [PubMed]
46. Kellner, F.; Geu-Flores, F.; Sherden, N.H.; Brown, S.; Foureau, E.; Courdavault, V.; O'Connor, S.E. Discovery of a P450-catalyzed step in vindoline biosynthesis: A link between the aspidosperma and eburnamine alkaloids. *Chem. Commun.* **2015**, *51*, 7626–7628. [CrossRef]
47. Liscombe, D.K.; Usera, A.R.; O'Connor, S.E. Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 18793–18798. [CrossRef]
48. Vazquez-Flota, F.; De Carolis, E.; Alarco, A.-M.; De Luca, V. Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in Catharanthus roseus (L.) G. Don. *Plant Mol. Biol.* **1997**, *34*, 935–948. [CrossRef] [PubMed]
49. De Carolis, E.; De Luca, V. A novel 2-oxoglutarate-dependent dioxygenase involved in vindoline biosynthesis: Characterization, purification and kinetic properties. *Plant Cell Tissue Organ Cult. (PCTOC)* **1994**, *38*, 281–287. [CrossRef]
50. Rabeharindranto, H.; Cerezo, S.C.; Lautier, T.; Garcia-Alles, L.; Treitz, C.; Tholey, A.; Truan, G. Enzyme-fusion strategies for redirecting and improving carotenoid synthesis in S. cerevisiae. *Metab. Eng. Commun.* **2019**, *8*, e00086. [CrossRef] [PubMed]
51. Kulagina, N.; Besseau, S.; Papon, N.; Courdavault, V. Peroxisomes: A New Hub for Metabolic Engineering in Yeast. *Front. Bioeng. Biotechnol.* **2021**, *9*, 659431. [CrossRef]
52. Zhu, J.; Schwartz, C.; Wheeldon, I. Controlled intracellular trafficking alleviates an expression bottleneck in S. cerevisiae ester biosynthesis. *Metab. Eng. Commun.* **2019**, *8*, e00085. [CrossRef] [PubMed]
53. Thodey, K.; Galanie, S.; Smolke, C.D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **2014**, *10*, 837–844. [CrossRef]
54. Stander, E.A.; Sepúlveda, L.J.; De Bernonville, T.D.; Carqueijeiro, I.; Koudounas, K.; Cruz, P.L.; Besseau, S.; LaNoue, A.; Papon, N.; Giglioli-Guivarc'H, N.; et al. Identifying Genes Involved in alkaloid Biosynthesis in *Vinca minor* Through Transcriptomics and Gene Co-Expression Analysis. *Biomolecules* **2020**, *10*, 1595. [CrossRef]
55. van Dijken, J.; Bauer, J.; Brambilla, L.; Duboc, P.; Francois, J.; Gancedo, C.; Giuseppin, M.; Heijnen, J.; Hoare, M.; Lange, H.; et al. An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains. *Enzym. Microb. Technol.* **2000**, *26*, 706–714. [CrossRef]
56. Urban, P.; Mignotte, C.; Kazmaier, M.; Delorme, F.; Pompon, D. Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related Arabidopsis thalianaNADPH-Cytochrome P450 Reductases with P450 CYP73A5. *J. Biol. Chem.* **1997**, *272*, 19176–19186. [CrossRef] [PubMed]
57. Suga, M.; Isobe, M.; Hatakeyama, T. Cryopreservation of competent intact yeast cells for efficient electroporation. *Yeast* **2000**, *16*, 889–896. [CrossRef]
58. Gietz, R.D.; Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2007**, *2*, 31–34. [CrossRef]

59. Parage, C.; Foureau, E.; Kellner, F.; Burlat, V.; Mahroug, S.; Lanoue, A.; de Bernonville, T.D.; Londono, M.A.; Carqueijeiro, I.; Oudin, A.; et al. Class II Cytochrome P450 Reductase Governs the Biosynthesis of Alkaloids. *Plant Physiol.* **2016**, *172*, 1563–1577. [CrossRef] [PubMed]
60. Rabin, R.C. Faced with a Drug Shortfall, Doctors Scramble to Treat. Children with Cancer. *New York Times*, 14 October 2019; Section A. 21.
61. Courdavault, V.; O'Connor, S.E.; Jensen, M.K.; Papon, N. Metabolic engineering for plant natural products biosynthesis: New procedures, concrete achievements and remaining limits. *Nat. Prod. Rep.* **2021**. [CrossRef]

Supplementary data:

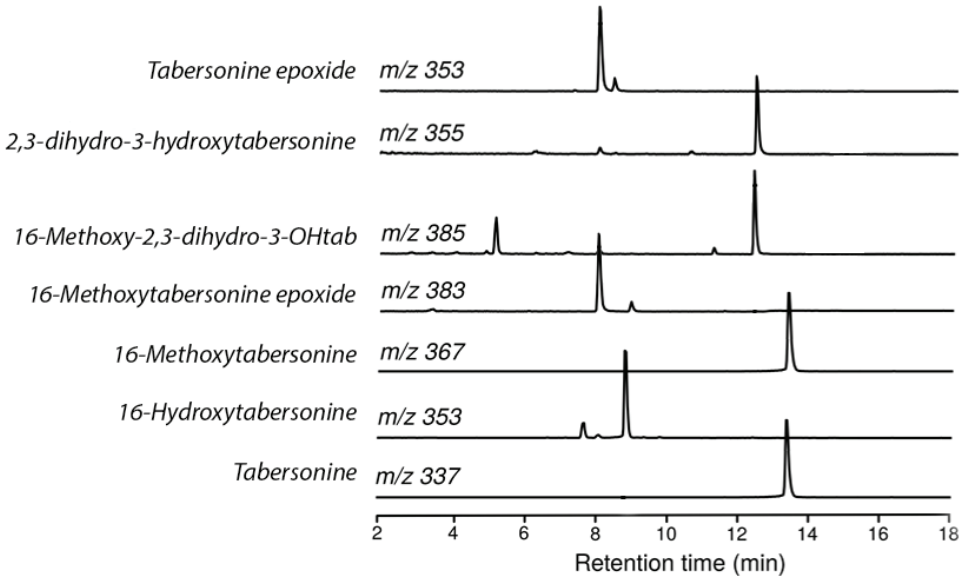


Figure 1. UPLC-MS chromatograms of the natural products produced by the yeast strains.

Table S1: Primers list

Gene	Forward primer	Reverse primer
T16H2	CTGAGAACTAGTTCGGTCATGGAGTTGTATTATTTTTCCACC	CTGAGAACTAGTCTAATATTTACCTTTGAGAGAAGAAGC
16OMT	CTGAGAACTAGTTCGGTCATGGATGTTCAATCTGAGGAGT	CTGAGAACTAGTTCAAGGATAAACCTCAATGAGAC
T3O	CTGAGAACTAGTTCGGTCATGGAGTTTCATGAATCTTCTCCC	CTGAGAACTAGTTCATGCATAGGACGTAGCGA
T3R	CTGAGATCTAGATCGGTCATGGCAATGGCTTCAAAGTC	CTGAGATCTAGATTAGGGTGATTTGAAAGTGTTTCC
T16H2_helix	CTGAGAACTAGTATGGAGTTGTATTATTTTTCCACCTTTG	GCCAGATTTCTTTAGAGTTTTGGCTAA
EROMT (fusion primer)	ACTCTAAAGAAATCTGGCATGGATGTTCAATCTGAGGAGTTCC	CTGAGAACTAGTTCAAGGATAAACCTCAATGAGACTCC
pTEF1	CCAGTCGATTATCATGTTCGCTACCACACACCATAGCTTCAAAATG	CTGAGAACTAGTTTTGTAATTAAAACTTAGATTAGATTGCTATGC
pPGK1	TAGCGAACATGATAATCGACTGGAGACGCGAATTTTTCGAAGAA	CTGAGAACTAGTTCTAGATGTTTTATATTTGTTGTAAAAAGTAGATAATTAC
pTDH3	TAGCGAACATGATAATCGACTGGCTATTTTCGAGGACCTTGTCACCTT	CTGAGAACTAGTTCTAGATTTGTTTGTTTATGTGTGTTTATTCGAAAC

pACT1	CCAGTCGATTATCATGTTCGCTAGTGAAGATGTGGCTGC AAGATT	CTGAGATCTAGAGCGGCCGCGTTAATTCAGTAAA TTTTCGATCTTGG
CPR *C. roseus*	GGATCCCCCTCGAGTTAATTAACGTAATGGATTCTTCCT CAGAAAAGTTAAGC	GAATTCGCTAGCTCTAGAGCGATCGCCCACTTAC CAAACATCACGTAAGTATCTACC
ARG3	CATTACGCTCCTTCGTATTAC	GTGAAATTGGACAACTCGAAGG
tCPS1	GAGCTCTTTAGTCATTGTATGGTC	ACTAGTGCGCAATGATTGAATAGTCAAAG
HIS5	TCTAGAATAGATTAATTTAAACAGTATATGTACAG	GTCGACTGATATTTAGCTATATGTACGTTGTTAG
URA3	GATTCGGTAATCTCCGAAC	TTTGTGAGTTTAGTATACATGC
IDP1	GAGCTCCAACAATAAGGTATATATATTTATGATAACAG	ACTAGTTCGAATTTACGTAGCCCAATC
PRM5	TCTAGAAAACTTTTATGATATTTTGCAATATTTTTTTTAA GC	CTCGAGTATAATAAGACACGGACGCAC
SAM2	CTGAGAACTAGTATGTCCAAGAGCAAAACTTTCTTATTT ACC	CTGAGAACTAGTTTAAAATTCCAATTTCTTTGGTT TTTCC



Figure S2: Fusion of the T16H2 transmembrane helix to the N-terminal end of 16OMT (ER_16OMT). Alignment of the first 55 residues of T16H2 with ER_16OMT. The red rectangle highlights the added sequence including the predicted transmembrane helix of identified in T16H2.

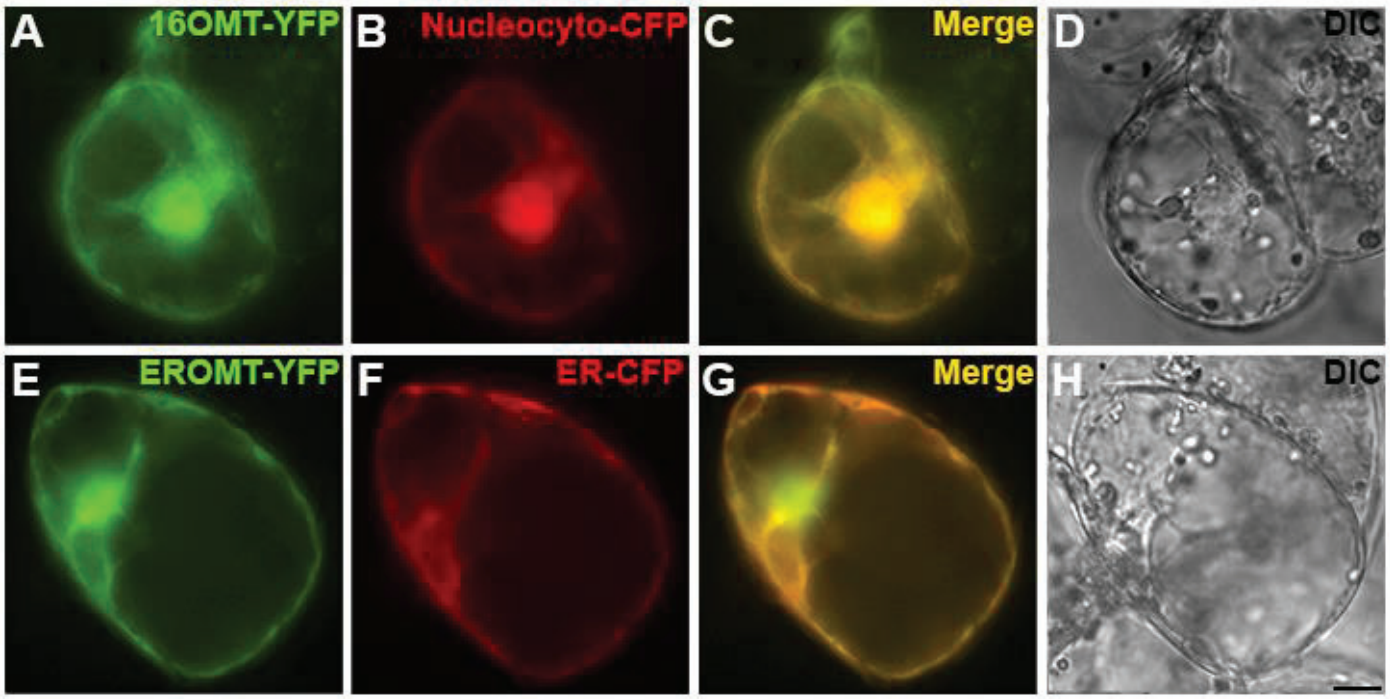


Figure S3: Subcellular localization of 16OMT and EROMT in *C. roseus* cells. Cells were transformed transiently with 16OMT-YFP (A–D) and EROMT-YFP (E-H) expressing vectors in combination with CFP-nucleocytosolic or CFP-ER marker (second column). Co-localization of the two fluorescence signals appeared in the merged image (C, G). The morphology is observed with differential interference contrast (DIC). Bar: 10 um.

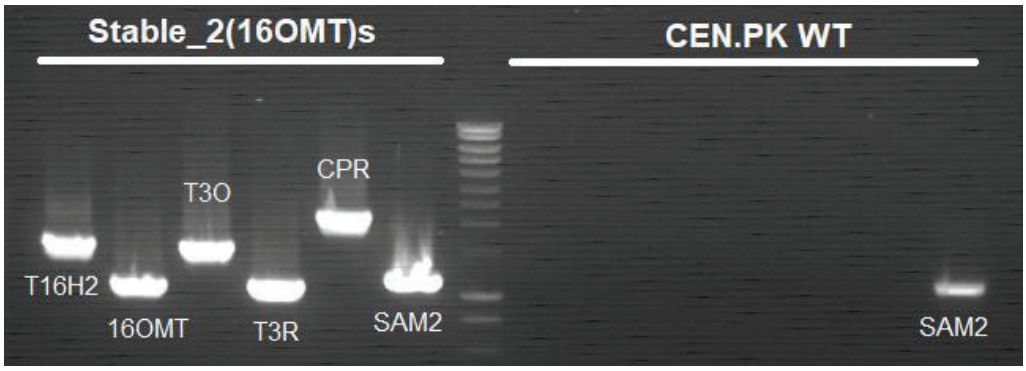


Figure S4: Phusion PCR amplification of the integrated genes from the vindoline's pathway using genomic DNA from the designed stable yeast (Stable_2(16OMT)s) and genomic DNA from wild type CEN.PK (CEN.PK WT). The yeast gene SAM2 (S-adenosylmethionine synthetase) was used as positive control. T16H2: tabersonyne-16-hydroxylase, 16OMT: tabersonine-16-*O*-methyltransferase, T3O: tabersonine 3-oxygenase, T3R: tabersonine 3-reductase, CPR: optimized *C. roseus CPR*, SAM2: S-adenosylmethionine synthetase.

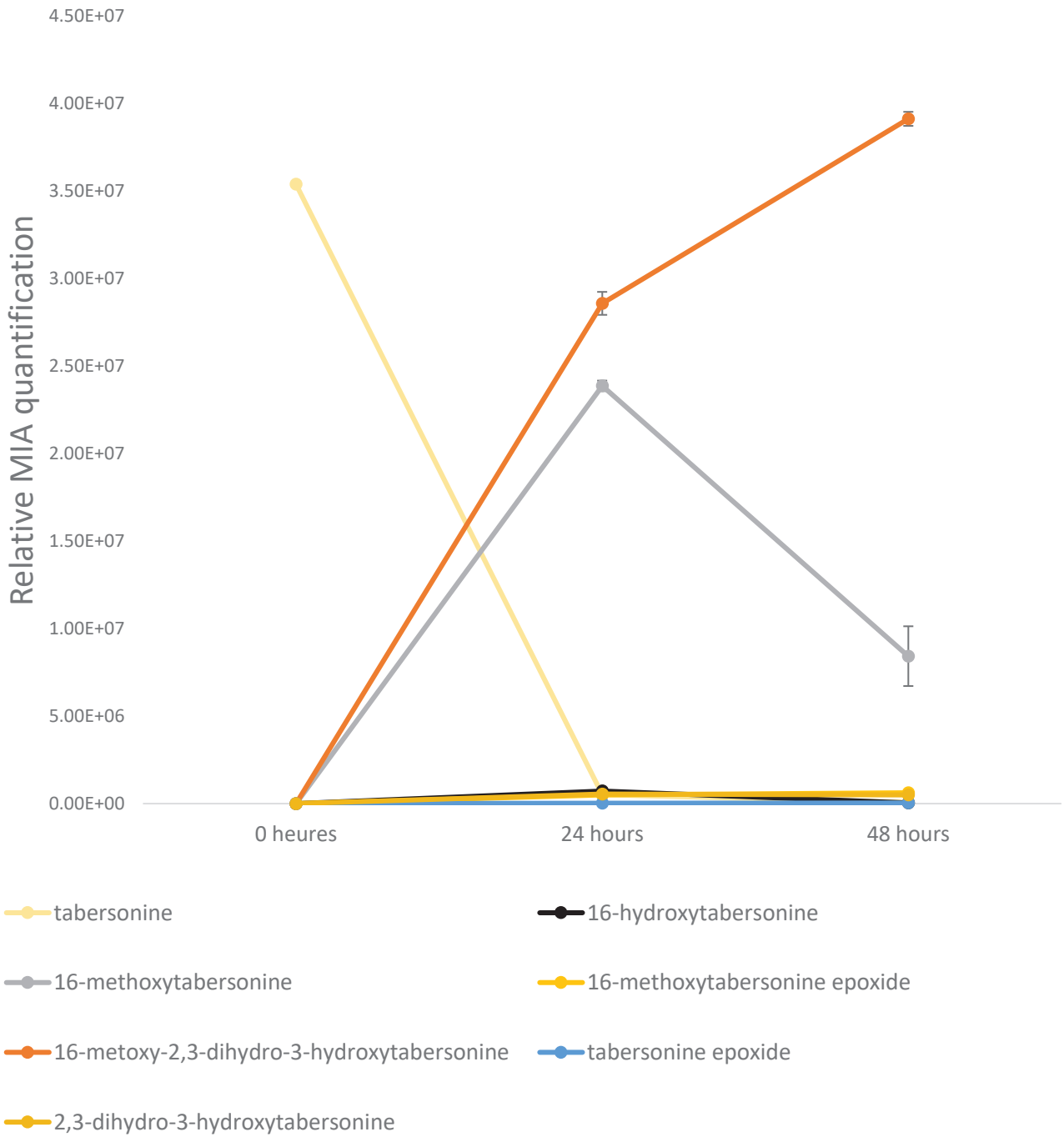


Figure S5: Evolution of the accumulation of vindoline and vindorosine biosynthetic intermediates in the stable_2(16OMT)s yeast strain fed with tabersonine. Alkaloids were quantified by UPLC-MS in the yeast culture medium before and 24 and 48 hours post-feeding with tabersonine (250 μM). Error bars correspond to the standard error of biological replicates (n = 3).

Article 6 : Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories

In the article 6, I participated in the metabolic engineering of the yeast strains (integration of the vindoline pathway genes) and on the bioconversion assays. The design on the integration plasmid backbones and bioreactor conducted experiments were performed by other authors.

RESUME EN FRANÇAIS DE L'ARTICLE 6

L'industrie pharmaceutique fait face à une demande croissante et à des pénuries récurrentes de nombreux métabolites d'origine végétale possédant des propriétés anticancéreuses et largement utilisées en chimiothérapie humaine. Des stratégies alternatives et efficaces d'approvisionnement de ces produits naturels tels que la bioproduction par des micro-organismes sont nécessaires pour assurer une production importante et pérenne. Nous avons développé et optimisé des levures utilisées comme cellules-usines convertissant efficacement la tabersonine en vindoline, un précurseur des principaux alcaloïdes anticancéreux vinblastine et vincristine. Ainsi, le réglage fin du nombre de copies de gènes hétérologues circonscrit la synthèse à la production de vindoline. La bioconversion de tabersonine en vindoline a été améliorée grâce à une optimisation rationnelle du milieu (pH, composition) et des conditions de culture via une stratégie d'alimentation séquentielle de substrats nutritifs. Finalement, une production de 266 $mg.l^{-1}$ (rendement 88 %) de vindoline a été atteinte en bioréacteur en mode discontinu alimenté (fed-batch). Cette synthèse de vindoline ouvre la voie à une future bioproduction industrielle à travers la valorisation d'abondantes ressources de tabersonine.

Check for updates

microbial biotechnology

Open Access

Brief Report

Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories

Natalja Kulagina,[1,†] Grégory Guirimand,[1,2,3,†] Céline Melin,[1,†] Pamela Lemos-Cruz,[1] Ines Carqueijeiro,[1] Johan-Owen De Craene,[1] Audrey Oudin,[1] Vladimir Heredia,[1] Konstantinos Koudounas,[1] Marianne Unlubayir,[1] Arnaud Lanoue,[1] Nadine Imbault,[1] Benoit St-Pierre,[1] Nicolas Papon,[4] Marc Clastre,[1] Nathalie Giglioli-Guivarc'h,[1] Jillian Marc,[1] Sébastien Besseau[1,*] and Vincent Courdavault[1,*]

[1]*Université de Tours, EA2106 Biomolécules et Biotechnologies Végétales, Tours, France.*
[2]*Graduate School of Sciences, Technology and Innovation, Kobe University, Kobe, Japan.*
[3]*Le Studium Loire Valley Institute for Advanced Studies, Orléans & Tours, France.*
[4]*Université d'Angers, EA3142 Groupe d'Etude des Interactions Hôte-Pathogène, Angers, France.*

Summary

The pharmaceutical industry faces a growing demand and recurrent shortages in many anticancer plant drugs given their extensive use in human chemotherapy. Efficient alternative strategies of supply of these natural products such as bioproduction by microorganisms are needed to ensure stable and massive manufacturing. Here, we developed and optimized yeast cell factories efficiently converting tabersonine to vindoline, a precursor of the major anticancer alkaloids vinblastine and vincristine. First, fine-tuning of heterologous gene copies restrained side metabolites synthesis towards vindoline production. Tabersonine to vindoline bioconversion was further enhanced through a rational medium optimization (pH, composition) and a sequential feeding strategy. Finally, a vindoline titre of 266 mg l^{-1} (88% yield) was reached in an optimized fed-batch bioreactor. This precursor-directed synthesis of vindoline thus paves the way towards future industrial bioproduction through the valorization of abundant tabersonine resources.

Received 14 April, 2021; accepted 10 July, 2021.
For correspondence. *E-mail sebastien.besseau@univ-tours.fr. **E-mail vincent.courdavault@univ-tours.fr; Tel. +33.2.47.36.70.23; Fax +33.2.47.36.72.32.
[†]Equal contribution.
Microbial Biotechnology (2021) **0**(0), 1–7
doi:10.1111/1751-7915.13898

Introduction

Vinblastine and vincristine are two major anticancer monoterpene indole alkaloids (MIAs) accumulated in the leaves of *Catharanthus roseus* and extensively used in human chemotherapies. Due to a constantly growing demand, both vinblastine and vincristine are suffering from recurrent shortages caused by socio-economic (e.g. demand pressure) and/or geopolitical and climatic constraints (Rabin, 2019). Additionally, the intensive culture of *C. roseus* conducted for the supply of these two drugs has a dramatic negative impact on the environment (e.g. water resource, pollution...) driving to the necessity of alternative methods for more sustainable production according to the Sustainable Development Goals (SDGs) of the United Nations (https://www.un.org/sustainabledevelopment/sustainable-development-goals). In such a perspective, the bioproduction of MIAs in heterologous hosts constitutes an attractive option to ensure a stable supply to the pharmaceutical industry (Courdavault *et al.*, 2020). Among all generally regarded as safe (GRAS) microorganisms, *Saccharomyces cerevisiae* (baker yeast) displays many advantages such as a well-known sequenced genome, as well as numerous molecular tools for genetic manipulation and constitutes a highly valuable platform for industrial bioproduction of plant natural products (Guirimand *et al.*, 2020). Since both vinblastine and vincristine derive from the condensation of catharanthine and vindoline, a stable supply of these two anticancer compounds could rely on the development of yeast cell factories by metabolic engineering of *S. cerevisiae* to produce vindoline and catharanthine as precursors. Conceptually, the vindoline synthesis can be performed either *de novo* (i.e. from glucose as original carbon source) or through precursor-directed synthesis by feeding engineered yeast with tabersonine, extractible at multigram scale from the

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

highly abundant seeds of *Voacanga africana* (Kikura-Hanajiri *et al*., 2009; Koroch *et al*., 2009). In this case, seven enzymatic steps are required for the biotransformation of tabersonine (Fig. 1A; Qu *et al*., 2015). Although the bioproduction of vindoline by yeast cell factories has been previously reported, the synthesis remains at a scarce level and thus requires optimization to reach industrial scale (Qu *et al*., 2015).

In the present study, the tabersonine-to-vindoline biosynthetic pathway from *C. roseus* was stably integrated within hotspots of the genome of *S. cerevisiae*, using CRISPR-Cas9 molecular tools (Mikkelsen *et al*., 2012). The fine-tuning of the copy numbers of the pathway genes and the optimization of the fermentation conditions allowed us to synthesize an unprecedented amount of vindoline, thus paving the way for a complementary production process of anticancer MIAs through metabolic engineering combined with the valorization of a renewable natural resource.

Results and discussion

Engineering of a vindoline-producing yeast strain

The bioconversion of tabersonine to vindoline requires seven enzymatic steps starting with T16H hydroxylating tabersonine (1) at carbon 16 to form 16-hydroxytabersonine (2), which is further methylated by 16OMT to 16-methoxytabersonine (3) Downstream conversion to 16-methoxytabersonine epoxide (4) is catalysed by T3O, which in turn is reduced to 16-methoxy-2,3-dihydro-3-hydroxytabersonine (5) by T3R. In the following steps, the T3R product is converted to desacetoxyvindoline (6) by NMT, then hydroxylated to deacetylvindoline (7) by D4H and finally converted to vindoline (8) by DAT (Fig. 1A; Courdavault *et al*., 2014). However, substrate promiscuity enables T3O using tabersonine directly to form tabersonine imine alcohol (9) (Kellner *et al*., 2015). This marks the onset of a parallel branch pathway leading to the production of undesirable by-products including 2,3-dihydro-3-hydroxytabersonine (10), desacetoxyvindorosine (11), deacetylvindorosine (12) and vindorosine (13), whose *in planta* amounts depend on T16H expression level (Besseau *et al*., 2013). Consequently, the heterologous production of vindoline may face suboptimal flux according to the initial decoration of tabersonine by T16H or T3O.

In order to create yeast cell factories producing vindoline from tabersonine, we selected the prototrophic *S. cerevisiae* strain CEN.PK113-7D, commonly used in metabolic engineering for industrial purposes. We initially implemented the five first genes of the tabersonine-to-vindoline biosynthetic pathway to produce desacetoxyvindoline through tabersonine transformation (strain T3110). T16H2, 16OMT, T3O, T3R, NMT coding sequences and ATR2 (required for T16H and T3O reduction) were integrated into three distinct hotspots of the *S. cerevisiae* genome, selected for high transcription activity (Mikkelsen *et al*., 2012), under strong glycolytic promoters (Table S2). To assess the production of desacetoxyvindoline, the T3110 strain was fed with 125 μM of tabersonine and grown for 120 h. Culture medium was collected every 24 h, and extracellular metabolite content was analysed by liquid chromatography coupled with mass spectrometry (LC-MS) since the majority of alkaloids are accumulated in the extracellular medium (Qu *et al*., 2015; Fig. S1A). Although tabersonine was not fully consumed after 24 h, the T3110 strain produced both desacetoxyvindoline and desacetoxyvindorosine in similar proportions (Fig. 1B). This confirms that the metabolic flux is simultaneously directed towards vindoline and vindorosine pathways as a result of T16H and T3O competition for tabersonine consumption. Therefore, in order to limit the vindorosine accumulation and improve the efficiency of tabersonine bioconversion, one additional copy of both T16H and 16OMT genes was introduced into the T3110 strain. The resulting daughter strain, named Q7, showed an 11-fold decrease of desacetoxyvindorosine accumulation, while desacetoxyvindoline drastically increased and the tabersonine was almost completely consumed after 24 h. Due to this improved metabolism and given that further increase in T16H/16OMT gene copy number did not impact desacetoxyvindoline production (Q112 strain), the Q7 strain was used as a background for the downstream integration of the final vindoline biosynthesis steps D4H and DAT. The resulting P12 strain, containing a single copy of D4H and DAT, produced 14.7 mg l^{-1} of vindoline (Table S3). However, an important accumulation of desacetoxyvindoline was also observed, highlighting a D4H-dependent bottleneck. To circumvent this lack of activity, a total of three copies of both D4H and DAT genes were introduced into the Q7 strain to generate the Hep3 strain. As a consequence, the addition of two supplemental copies of D4H and DAT increased vindoline production up to 26.8 mg l^{-1} after 120 h (i.e. two times higher than P12 strain) and halved the accumulation of desacetoxyvindoline. Interestingly, adding a fourth copy of both D4H and DAT genes did not improve vindoline titre (Hep122 strain), suggesting that other parameters such as culture conditions could be limiting rather than the number of gene copies. Moreover, after 72–96 h the bioconversion was reaching a plateau, which is probably due to the depletion of nutrients in the medium and an increasing competition with native metabolism for cofactors. Thus, Hep3 strain was selected for further optimization of vindoline production.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

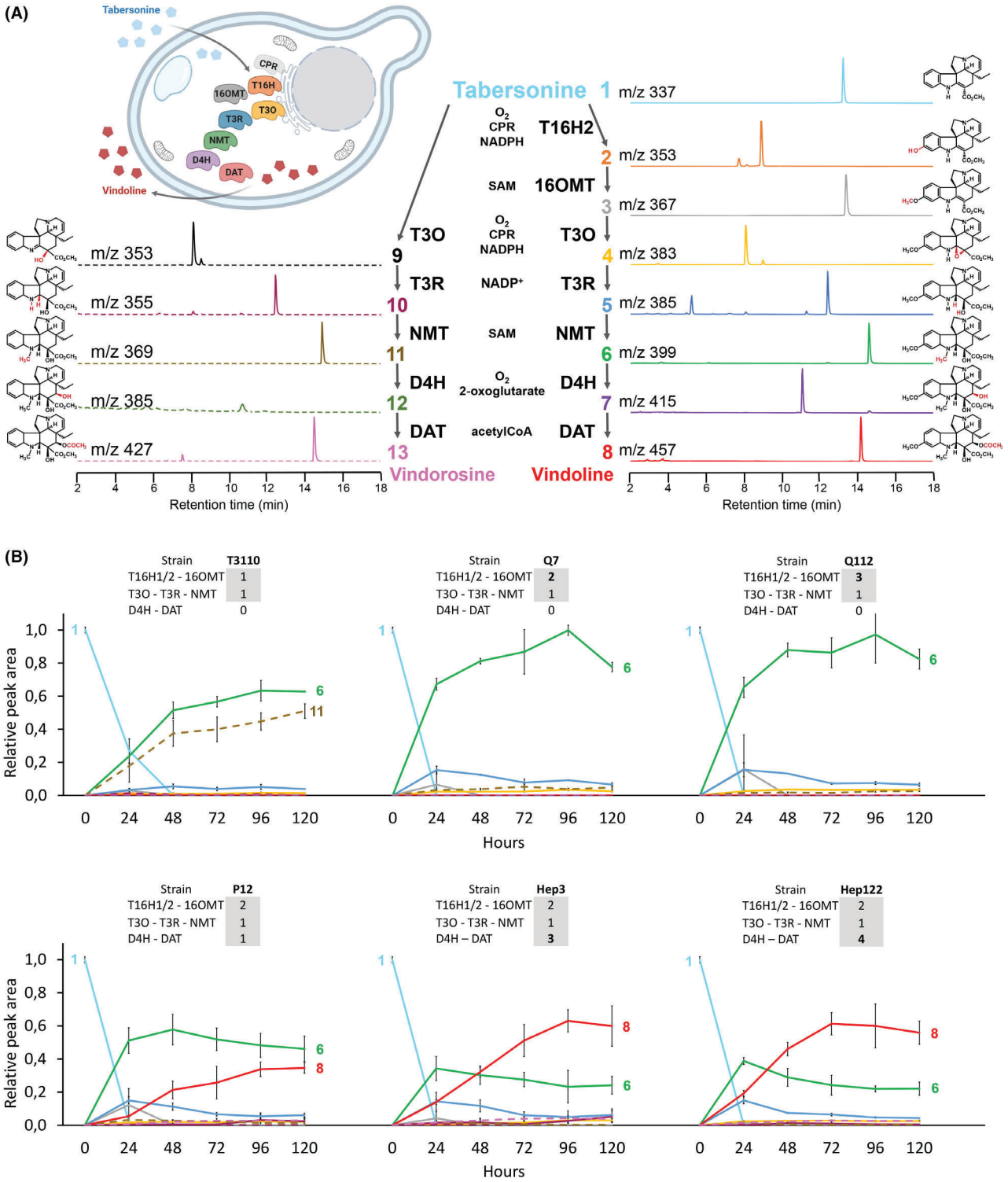


Fig. 1. Tailoring yeast cell factories for vindoline bioproduction.
A. Vindoline biosynthetic pathway and parallel branch vindorosine pathway. Each colour and number correspond to enzyme product, represented by molecular structure, chromatogram, m/z and retention time: 1 tabersonine, 2 16-hydroxytabersonine, 3 16-methoxytabersonine, 4 16-methoxytabersonine epoxide, 5 16-methoxy-2,3-dihydro-3-hydroxytabersonine, 6 desacetoxyvindoline, 7 deacetylvindoline, 8 vindoline and the by-products 9 tabersonine imine alcohol, 10 2,3-dihydro-3-hydroxytabersonine, 11 desacetoxyvindorosine, 12 deacetylvindorosine and 13 vindorosine. MS/MS fragmentation patterns of compounds are presented in Table S1. Cofactors are indicated for each enzyme.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

Fig 1. (Continued)

B. Time-course monitoring of extracellular metabolite content in generated *S. cerevisiae* strains. The feedings were performed with 125 μM of tabersonine in the initial 200 μl of YPD/strain/time point. The names and the number of gene copies integrated into each strain are written on the top of each graphic. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation (n = 3 biological replicates). By-products are shown in dashed lines.

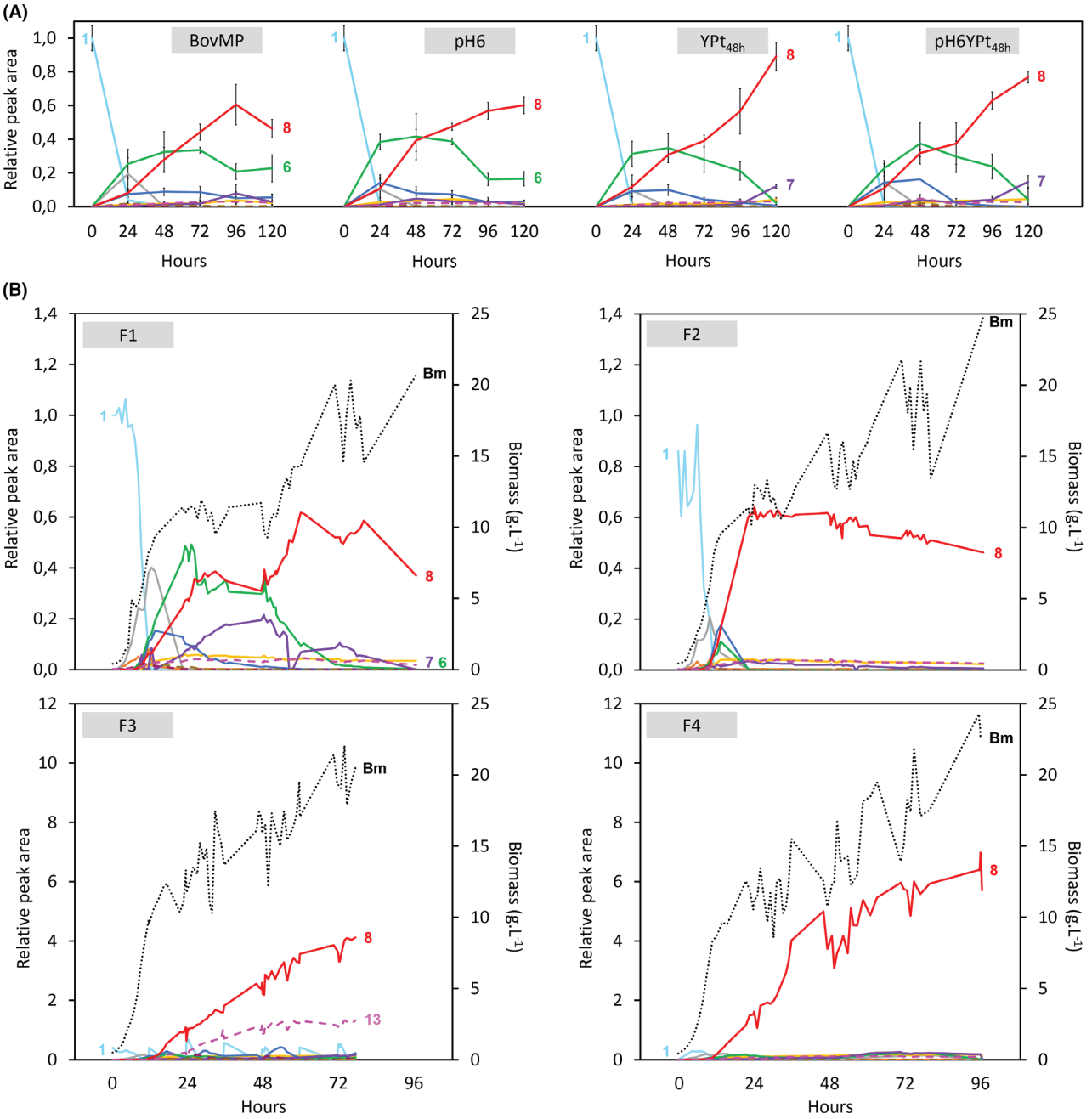


Fig. 2. Optimization of vindoline production at small and medium scales.
A. Effect of pH and addition of fresh YP on small-scale vindoline production. The tested condition is notified on the top of each graphic. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation (n = 3 biological replicates). The feedings were performed with 125 μM of tabersonine in the initial 200 μl of YPD/strain/time point.
B. Effect of growth substrate and tabersonine feeding strategies on vindoline production in fed-batch cultivation. Growth substrates are fed by pulses (F1) or continuously (F2-F4). Tabersonine is fed by one pulse (F1, F2), seven pulses (F3) or equivalent continuous feeding (F4) (Table S4). Bm, biomass. The replicate of F4 feeding strategy is shown in Fig. S3A.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

Table 1. Main kinetic parameters measured during vindoline production in the bioreactor.

Strategy	Maximal growth rate (h^{-1})	Max concentrations ± SD		Conversion yield (% of theoretical maximal yield) (g .g^{-1})	Productivity (mg. l^{-1} .h^{-1})	Final fraction of vindoline on end-products (g. g^{-1})
		Vindoline (mg. l^{-1})	Vindorosine (mg. l^{-1})			
F0	0.36	–	–	–	–	–
F1	0.30	19.1 ± 0.3	1.4 ± 0.1	0.42 (31%)	0.32	0.92
F2	0.30	19.2 ± 0.5	1.5 ± 0.1	0.43 (32%)	1.7	0.94
F3	0.29	161 ± 2	51 ± 2	0.62 (45%)	2.1	0.75
F4	0.31	266 ± 8	4.7 ± 0.4	1.20 ± 0.01 (88%)	2.77 ± 0.02	0.98

Growth substrates are fed by pulses (F1) or continuously (F2-F4). Tabersonine is fed by one pulse (F1, F2), seven pulses (F3) or equivalent continuous feeding (F4) (Table S4). Standard deviation: $n = 4$ independent samples, considering the duplicate experiment for F4. End products are vindoline and vindorosine.

Optimization of culture conditions to enhance vindoline production

In order to evaluate the potential effect of culture medium and nutrient supply on the production of vindoline by the vindoline-producing strain Hep3, several sources of peptone nutrients (Table S3) were assayed. Compared with the bovine peptone (BovMP) routinely used, another peptone produced from bovine tissues (BovP), a peptone obtained from a mix of bovine and porcine tissues (BovPrcP), as well as tryptone and polypeptone (PolyP), showed drastic negative effects on the activity of NMT and D4H leading to the accumulation of intermediates 5 and 6, along with the poor production of vindoline (Fig. S1B). Consequently, BovMP was selected for further optimization of culture conditions. Above all, these results demonstrate that growth media composition has an acute effect on bioproduction underlining the importance of a careful choice of peptone sources.

Similarly, the pH of the culture medium was found to drastically impact the production of vindoline, pH 6.0 being determined as optimal (Fig. S1C). As the acidification of 2 points of the culture medium pH was observed after 120 h (Table S3), the importance of maintaining a constant level of optimal pH was investigated. In parallel, to prevent any starvation-related metabolic stress, the supply with fresh YP nutrients (yeast extract and peptone) every 24h after 48h of culture was assessed. The best vindoline production of 34.6 mg l^{-1-1} was achieved by adding YP (condition YPt_{48h}), which demonstrated an enhanced activity of D4H after 96h of culture and a 1.5-fold increase in vindoline titre compared with standard conditions (condition BovMP) (Fig. 2A, Table S3). Interestingly, the sole addition of YP was sufficient to maintain pH close to 6.0, driving to a slightly higher vindoline production compared with additional buffering of culture medium (condition $pH6YPt_{48h}$). Hence, using BovMP peptone as the best source of nutrients and supplementing the culture with YP every 24 h from 48 h post feeding was defined as the optimal conditions for vindoline production.

Scaling up and engineering of bioreactor feeding to improve vindoline production

In order to scale up the production of vindoline, the Hep3 strain was next cultivated in bioreactors. In control conditions, both wild-type and Hep3 strains showed similar maximal growth rates (0.37 and 0.36 h^{-1} respectively), viability (maintained up to 98%), as well as biomass and ethanol yields on glucose. This tends to show that overexpression of the vindoline pathway in the recombinant strain was well tolerated and did not constitute any utter metabolic burden for the cells. However, the addition of tabersonine into the culture medium increased the cellular respiration on oxygen, while the maximal growth rate showed a 20% decrease (Table 1). Such phenomenon, routinely observed with the addition of weak acid (Verduyn *et al*., 1992), reveals an uncoupling effect of tabersonine on energy and biomass productions that did not alter cell viability in all of the experiments (up to 98% in all the cultivations) (Fig. S2).

The F1 feeding experiment in the bioreactor was based on the best production strategy previously defined, relying on 24 h-pulsated supplementations of glucose followed by YPD addition. In these conditions, the conversion yield was found to be two times lower than in small-scale experiments (Table 1). However, the speed of production of vindoline was two times higher, with a maximal amount of vindoline produced within only 60 h, while the competing flux leading to the formation of vindorosine remained low (Fig. 2B; Fig. S2). In addition, except for the last stage of the starting pulse, a substantial decrease in vindoline cumulated mass was observed when glucose and ethanol were depleted, thus suggesting its consumption (Fig. S2). Therefore, in order to improve vindoline production, culture conditions were further optimized. For instance, glucose depletion between pulses was avoided by continuous feeding of growth substrates (F2 cultivation). Such feeding led to a 3 times faster conversion (22 h) than in the F1 reference culture, and consequently to higher productivity (Fig. 2B;

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

Table 1). In addition, vindoline was also continuously synthesized and not degraded any longer, even 3 days after the synthesis ended (Fig. S2). Moreover, while many biosynthetic intermediates were accumulated in the F1 condition, they were only detected in minute amounts after 22 h of bioconversion in F2, as a result of an extended MIA biosynthetic enzyme activity (Fig. 2B). Above all, this suggests that an active central primary metabolism is critical to provide sufficient amounts of cofactors required by the heterologous pathway, such as acetyl-coA and NADPH.

In the next step, we investigated the intensification of the vindoline production by numbering-up the pulses of tabersonine (F3 cultivation). Interestingly, growth and production were maintained all over the culture, and a *circa* 50% increase in the conversion yield was observed compared with F2 culture (Fig. 2B; Table 1). This marked increase probably resulted from cell environment management, reducing the impact of tabersonine uncoupling effect that induces huge pH and pO_2 drops. Nevertheless, the tabersonine flux towards vindorosine highly increased at the expanse of the vindoline pathway resulting in a 3:1 vindoline/vindorosine ratio (Table 1; Fig. 2B; Fig. S2). In addition, a higher accumulation of biosynthetic intermediates was also monitored during the whole experiment. To prevent the formation of undesired products and reduce heterologous pathway bottlenecks, we combined an uninterrupted active central metabolism (continuous glucose feeding under the Crabtree effect's critical uptake rate, avoiding ethanol production) with a restrained consumption rate of tabersonine (Fig. S2). Accordingly, we limited the specific uptake rate of tabersonine under its maximal value (1.03 ± 0.04 mg g^{-1} h^{-1}) through a continuous feeding in the F4 cultivation (Fig. S2; Fig. S3B). This strategy prevented tabersonine accumulation in the medium after 10 h and led to a 88% tabersonine bioconversion into vindoline, which was equivalently replicated (Fig. S3A). As a consequence, very low levels of vindorosine and vindoline biosynthetic intermediates were observed. (Table 1; Fig. 2B; Fig. S2; Fig. S3A). Thus, the control of central and heterologous metabolic fluxes by rational process design appeared as a relevant strategy to further improve vindoline production. In these conditions, vindoline reached a final titre of 266 mg l^{-1} after 96 h of culture, and thus, 20-fold higher productivity than the original biosynthesis reported by Qu et al. (2015) with an episomal strain. To date, it constitutes the highest vindoline production ever reported, in terms of yield, titre and productivity.

Conclusions

Through a fine-tuned integration of heterologous biosynthetic genes combined with the optimization of culture conditions, we generated a yeast strain synthesizing high amounts of vindoline *via* the bioconversion of tabersonine (88% yield). Based on the quantity of tabersonine accumulated in the seeds of *V. Africana* and annual exportation of 1600 tons of seeds, the total production of 22.5 tons of vindoline can be expected using this newly developed yeast strain (Kikura-Hanajiri *et al.*, 2009; Koroch *et al.*, 2009). This advance thus opens new perspectives towards the development of a potential alternative supply of vindoline to avoid vinblastine and vincristine shortage in the future.

Acknowledgements

We acknowledge funding from the ARD2020 Biopharmaceutical program of the Région Centre Val de Loire (BioPROPHARM, CatharSIS and ETOPOCentre projects), Le Studium Institute for a Research Consortium Grant and the EU Horizon 2020 research and innovation program (MIAMi project-grantagreement N○814645).

Funding information

We acknowledge funding from the ARD2020 Biopharmaceutical program of the Région Centre Val de Loire (BioPROPHARM, CatharSIS and ETOPOCentre projects), Le Studium Institute for a Research Consortium Grant and the EU Horizon 2020 research and innovation program (MIAMi project-grantagreement N○814645).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contribution

NK, GG, AO and JM contributed to conceptualization, methodology, investigation, data curation, writing – original draft and writing – review and editing. CM, PLC, IC and NP contributed to investigation and writing – review and editing. JDC VH, KK, MU and AL contributed to investigation. NI, BStP, MC and NGG contributed to formal analysis and investigation. SB and VC contributed to supervision, conceptualization, methodology, investigation, data curation, writing – original draft and writing – review and editing.

References

Besseau, S., Kellner, F., Lanoue, A., Thamm, A.M.K., Salim, V., Schneider, B., *et al.* (2013) A pair of tabersonine 16-

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant Physiol* **163:** 1792–1803.
Courdavault, V., O'Connor, S.E., Oudin, A., Besseau, S., and Papon, N. (2020) Towards the microbial production of plant-derived anticancer drugs. *Trends in Cancer* **6:** 444–448.
Courdavault, V., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B., and Burlat, V. (2014) A look inside an alkaloid multisite plant: the *Catharanthus logistics*. *Curr Opin Plant Biol* **19:** 43–50.
Guirimand, G., Kulagina, N., Papon, N., Hasunuma, T., and Courdavault, V. (2020) Innovative tools and strategies for optimizing yeast cell factories. *Trends Biotechnol* **39:** 488–504.
Kellner, F., Geu-Flores, F., Sherden, N.H., Brown, S., Foureau, E., Courdavault, V., and O'Connor, S.E. (2015) Discovery of a P450-catalyzed step in vindoline biosynthesis: a link between the aspidosperma and eburnamine alkaloids. *Chem Commun* **51:** 7626–7628.
Kikura-Hanajiri, R., Maruyama, T., Miyashita, A., and Goda, Y. (2009) Chemical and DNA analyses for the products of a psychoactive plant, Voacanga Africana. *Yakugaku Zasshi* **129:** 975–982.
Koroch, A.R., Juliani, H.R., Kulakowski, D., Arthur, H., Asante-Dartey, J., and Simon, J.E. (2009) Voacanga africana: chemistry, quality and pharmacological activity. *ACS Symp Ser* **1021:** 363–380.
Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H., and Halkier, B.A. (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab Eng* **14:** 104–111.
Qu, Y., Easson, M.L.A.E., Froese, J., Simionescu, R., Hudlicky, T., and DeLuca, V. (2015) Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc Natl Acad Sci USA* **112:** 6224–6229.
Rabin, R.C. (2019) *Faced With a Drug Shortfall, Doctors Scramble to Treat Children With Cancer*. New York Times. URL https://www.nytimes.com/2019/10/14/health/cancer-drug-shortage.html.
Sustainable Development Goals of The United Nations. URL https://www.un.org/sustainabledevelopment/sustainable-development-goals/.
Verduyn, C., Postma, E., Scheffers, W.A., and Van Dijken, J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8:** 501–517.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. (A) Extracellular and intracellular metabolite content in T3110 and Hep3 small-scale cultures 120 h post feeding. Supernatants (referred to as extracellular) were prepared as described in Experimental procedures of Appendix S1. Pellets were washed twice in 500 μl of saline solution (50 mM Tris, 100 mM NaCl), resuspended in 200 μl of MeOH, lysed by 10 min sonication, and the supernatants (referred to as intracellular) were further prepared as extracellular supernatants. The histograms represent the means of peak areas. Error bars: standard deviation ($n = 3$ biological replicates). (B) The effect of peptone on vindoline production and intermediate accumulation in Hep3 small-scale cultures. Hep3 feedings were set up in YPD media prepared with different peptones (Table S3). The feedings were performed with 125 μM of tabersonine in the initial 200 μl of YPD/strain/time point. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation ($n = 3$ biological replicates). (C) The effect of pH on extracellular metabolite content in Hep3 small-scale cultures. Hep3 feedings with tabersonine were set up in YPD medium buffered to pH 4, pH 5 or pH 6. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation ($n = 3$ biological replicates). The feedings were performed with 125 μM of tabersonine in the initial 200 μL of YPD/strain/time point.
Fig. S2. Evolution in bioreactors of (A) cumulated masses of tabersonine, vindoline, vindorosine and fraction of vindoline on end products, and evolution of (B) cumulated masses of glucose, biomass, ethanol and viability. Consumed tabersonine (dark square), vindoline (red diamond), vindorosine (pink triangle), fraction (yellow circle), consumed glucose (blue square), biomass (brown triangle), ethanol (green diamond) and viability (brown circle). F4 duplicate is in dotted lines and empty symbols. Growth substrates are fed by pulses (F1) or continuously (F2-F4). Tabersonine is fed by one pulse (F1, F2), seven pulses (F3) or equivalent continuous feeding (F4) (Table S4).
Fig. S3. (A) Replicate of the feeding strategy F4 presented in the Fig. 2B. Fresh medium and tabersonine are fed continuously (Table S4). (B) Evolution of the specific uptake rate of tabersonine in F4 bioreactor. Specific uptake rate (square), residual tabersonine (triangle) and biomass (circle). The replicate is in dotted line and empty symbols. The horizontal dotted line represents the maximal rate. Tabersonine is fed continuously by constant continuous feeding (Table S4).
Table S1. LC-MS/MS characterization of tabersonine and vindoline intermediates obtained with engineered yeasts. The fragmentation was performed on molecular ions $[M + H]+$ using cone voltage of 30 V and collision energies of 30 eV.
Table S2. Expression cassettes integrated in yeast hotspots.
Table S3. pH measurement and vindoline quantification in small-scale cultures.
Table S4. Strategies for bioreactor process optimization.
Table S5. List of primers used in the study.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd.

Supporting information

Enhanced bioproduction of anticancer precursor vindoline by yeast cell factories

Natalja Kulagina*[1], Grégory Guirimand*[1,2,3], Céline Melin*[1], Pamela Lemos-Cruz[1], Ines Carqueijeiro[1], Johan-Owen De Craene[1], Audrey Oudin[1], Vladimir Heredia[1], Konstantinos Koudounas[1], Marianne Unlubayir[1], Arnaud Lanoue[1], Nadine Imbault[1], Benoit St-Pierre[1], Nicolas Papon[4], Marc Clastre[1], Nathalie Giglioli-Guivarc'h[1], Jillian Marc[1], Sébastien Besseau#[1], Vincent Courdavault#[1]

1 Université de Tours, EA2106 Biomolécules et Biotechnologies Végétales, Tours, France

2 Graduate School of Sciences, Technology and Innovation, Kobe University, Kobe, Japan

3 Le Studium Loire Valley Institute for Advanced Studies, Orléans & Tours, France

4 Université d'Angers, EA3142 Groupe d'Etude des Interactions Hôte-Pathogène, Angers, France.

* equal contribution

#corresponding authors: sebastien.besseau@univ-tours.fr; vincent.courdavault@univ-tours.fr

Table S1: LC-MS/MS characterization of tabersonine and vindoline intermediates obtained with engineered yeasts. The fragmentation was performed on molecular ions [M + H]+ using cone voltage of 30 V and collision energies of 30 eV

	Compounds	Rt (min)	m/z ES+	Fragmentation	Reference
1	Tabersonine	13.5	337	305(27), 277(10), 249(23), 228(21), 196(33), 170(31), 168(100), 156(22), 144(17), 136(55), 122(4), 108(16)	Schweizer *et al.* 2018
2	16-hydroxytabersonine	9	353	321(44), 293(10), 265(41), 244(27), 212(70), 186(38), 184(100), 172(23), 160(17), 136(95), 122(44), 108(28)	Schweizer *et al.* 2018
3	16-methoxytabersonine	13.8	367	335(36), 307(11), 279(40), 258(24), 226(78), 200(39), 198(73), 186(23), 174(17), 136(100), 122(45), 108(19)	
4	16-methoxytabersonine epoxide	8	383	277(6), 240(6), 201(4), 174(100), 96(7)	
5	16-methoxy-2,3-dihydro-3-hydroxytabersonine	12.5	385	367(24), 337(10), 325(18), 307(100), 278(11), 200(12), 174(32), 147(25), 136(56), 122(36), 107(18)	
6	Desacetoxyvindoline	14.5	399	381(27), 339(28), 321(100), 282(13), 274(11), 272(18), 253(30), 246(9), 242(8), 214(18), 188(57), 186(24), 174(10), 122(77)	Colinas *et al*. 2021
7	Desacetylvindoline	11	415	397(1), 365(1), 355(17), 337(1), 295(4), 188(100), 180(4), 162(5), 122(9)	Colinas *et al*. 2021
8	Vindoline	14.2	457	439(1), 397(6), 365(1), 337(1), 222(2), 207(2), 194(1), 188(100), 179(1), 173(1), 162(2), 122(1)	Schweizer *et al.* 2018
13	Vindorosine	14.5	427	367(10), 355(4), 326(5), 307(4), 290(4), 265(3), 236(3), 223(3), 193(4), 158(100), 148(3), 122(2)	Colinas *et al*. 2021

Table S2: Expression cassettes integrated in yeast genome hotspots

Plasmid name	Expression cassettes	Integration locus
pPETA102-T16H2/ATR2	P_{TEF1}-T16H2-T_{ADH1} ; T_{PGK1}-ATR2-T_{ENO2}	X-2
pDPTA111-T16H1/16OMT	P_{TEF1}-T16H1-T_{ADH1} ; P_{TDH3}-16OMT-T_{PGI1}	XI-1
pPEAC115-NMT/T3R	P_{ACT1}-NMT-T_{CYC1} ; T_{PGK1}-T3R-T_{ENO2}	XI-5
pDPAC124-16OMT/T3O	P_{ACT1}-16OMT-T_{CYC1} ; P_{TDH3}-T3O-T_{PGI1}	XII-4
pPETA104- D4H/DAT	P_{TEF1}-D4H-T_{ADH1} ; T_{PGK1}-DAT-T_{ENO2}	X-4
pDPTA125- D4H/DAT	P_{TEF1}-D4H-T_{ADH1} ; P_{TDH3}-DAT-T_{PGI1}	XII-5
pDPAC113-D4H/DAT	P_{ACT1}-D4H-T_{CYC1} ; P_{TDH3}-DAT-T_{PGI1}	XI-3
pDPTA112-T16H1/16OMT	P_{TEF1}-T16H1-T_{ADH1} ; P_{TDH3}-16OMT-T_{PGI1}	XI-2
pDPAC122-D4H/DAT	P_{ACT1}-D4H-T_{CYC1} ; P_{TDH3}-DAT-T_{PGI1}	XII-2

Table S3: pH measurement and vindoline quantification in small-scale cultures

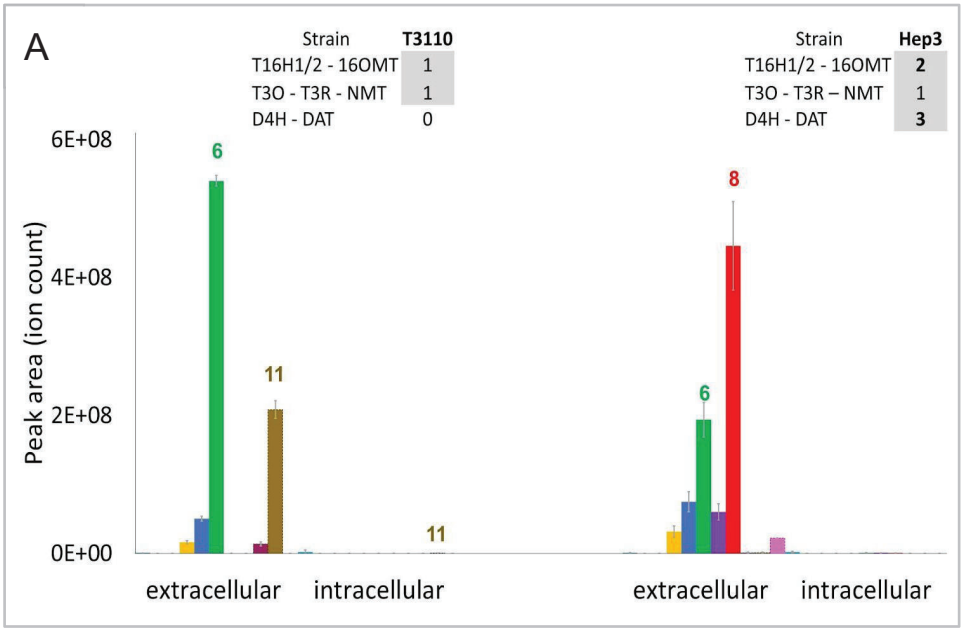
Vindoline - producing strains							
Strain	**0 h (pH)**	**24 h (pH)**	**48 h (pH)**	**72 h (pH)**	**96 h (pH)**	**120 h (pH)**	**Max concentration +/- SD ($mg.L^{-1}$)**
P12 Peptone from bovine muscles, peptic digestion, Sigma (70175)	-	-	-	-	-	-	14.7 +/-1.3
Hep3 Peptone from bovine muscles, peptic digestion, Sigma (70175)	-	-	-	-	-	-	26.8* +/- 2.9
Hep3 strain							
Condition	**0 h (pH)**	**24 h (pH)**	**48 h (pH)**	**72 h (pH)**	**96 h (pH)**	**120 h (pH)**	**Max concentration +/- SD ($mg.L^{-1}$)**
BovMP Peptone from bovine muscles, peptic digestion, Sigma (70175)	6	6	6	5-5.5	5	4	23.5 +/-4.7
BovP Peptone from bovine tissues, peptic digestion, Bio Basic (G213)	6	5.5-6	5	5	4-5	4	6.5 +/-2.3
BovPrcP Peptone from bovine and porcine tissues, peptic digestion, Thermo Fisher (2100677)	6	5.5	5	5	4	4	5.7 +/-3.4
Tryptone Animal peptone, casein digestion, Duchefa Biochemie (T1332)	6	6	5-6	5	5	4	14.5 +/-3
PolyP Animal peptone, peptic and casein digestion, Thermo Fisher (211910)	6	6	5.5	5	5	4	28.1 +/-4.6
pH6 (BovMP peptone)	6	6	6	6	6	6	23.3 +/-1.9
YPt_{48h} (BovMP peptone)	6	6	6	6	6	6	34.6** +/-3.2
$pH6YPt_{48h}$ (BovMP peptone)	6	6	6	6	6	6	29.8** +/- 1.3

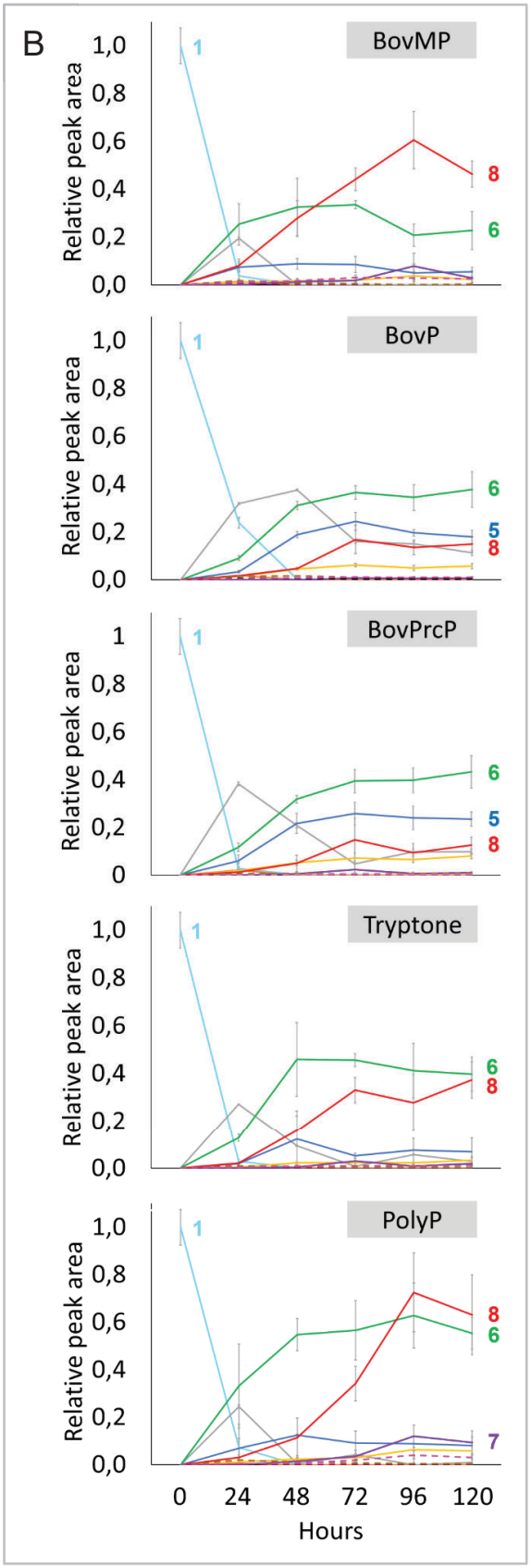
* The increase is significant compared to P12 strain with p-value < 0.05 (One-tailed T-test).

** The increase is significant compared to standard conditions (BovMP) with p-value < 0.05 (One-tailed T-test).

Table S4: Strategies for bioreactor process optimization

Strategy	Glucose	YPD	Tabersonine
F0	Pulsed at 20 $g.L^{-1}$ (t_{24h})	Pulsed at 20 $g.L^{-1}$ of glucose (t_{48h} and every 24h)	None
F1	Pulsed at 20 $g.L^{-1}$ (t_{24h})	Pulsed at 20 $g.L^{-1}$ of glucose (t_{48h} and every 24h)	Pulsed at 125 µM (1 pulse at t_0)
F2	Continuous at steady flow rate (t_0 to t_{48h}; equivalent to F1)	Continuous at steady flow rate (t_{48h} to the end; equivalent to F1)	Pulsed at 125 µM (1 pulse at t_0)
F3	Continuous at steady flow rate (t_0 to t_{48h}; equivalent to F1)	Continuous at steady flow rate (t_{48h} to the end; equivalent to F1)	Pulsed at 125 µM (1 pulse at t_0, 6 pulses every 12h)
F4	Continuous at steady flow rate (t_0 to t_{48h}; equivalent to F1)	Continuous at steady flow rate (t_{48h} to the end; equivalent to F1)	Continuous at steady flow rate (t_0 to the end time of tabersonine consumption in F3; equivalent to F3 pulses)


A
Strain T3110
T16H1/2 - 16OMT 1
T3O - T3R - NMT 1
D4H - DAT 0
Strain Hep3
T16H1/2 - 16OMT 2
T3O - T3R – NMT 1
D4H - DAT 3
Peak area (ion count)
6E+08
4E+08
2E+08
0E+00
extracellular
intracellular
extracellular
intracellular


B
BovMP
BovP
BovPrcP
Tryptone
PolyP
Relative peak area
Hours
0
24
48
72
96
120

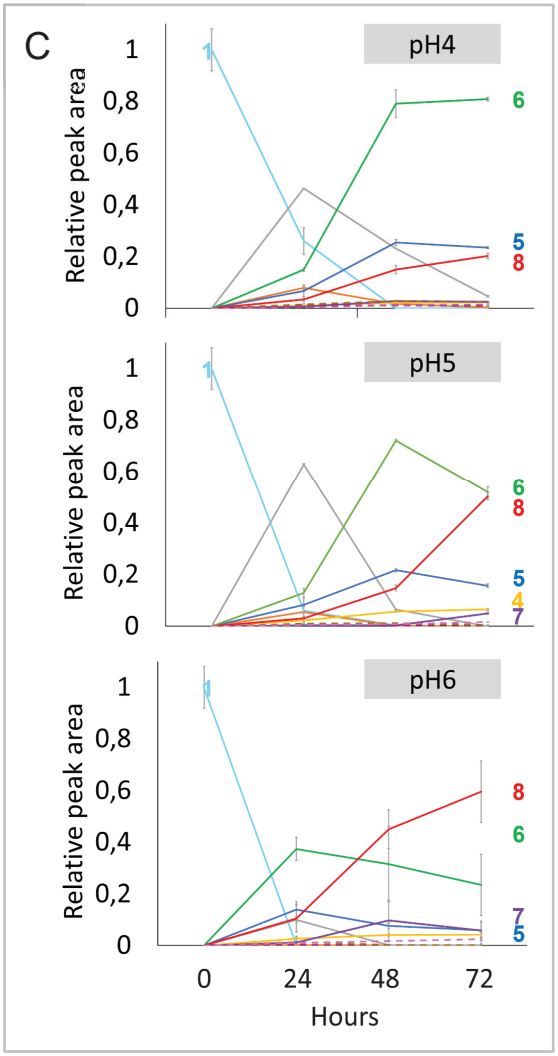

C
pH4
pH5
pH6
Relative peak area
Hours
0
24
48
72

Figure S1: (A) Extracellular and intracellular metabolite content in T3110 and Hep3 small-scale cultures 120 h post feeding. Supernatants (referred to as extracellular) were prepared as described in Experimental procedures of Appendix S1. Pellets were washed twice in 500µL of saline solution (50 mM Tris, 100 mM NaCl), resuspended in 200 µL of MeOH, lysed by 10 min sonication, and the supernatants (referred to as intracellular) were further prepared as extracellular supernatants. The histograms represent the means of peak areas. Error bars: standard deviation (n = 3 biological replicates). **(B) The effect of peptone on vindoline production and intermediate accumulation in Hep3 small-scale cultures.** Hep3 feedings were set up in YPD media prepared with different peptones (Table S3). The feedings were performed with 125 µM of tabersonine in the initial 200 µL of YPD/strain/time point. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation (n = 3 biological replicates). **(C) The effect of pH on extracellular metabolite content in Hep3 small-scale cultures.** Hep3 feedings with tabersonine were set up in YPD medium buffered to pH 4, pH 5 or pH 6. The curves represent the means of peak areas relative to tabersonine. Error bars: standard deviation (n = 3 biological replicates). The feedings were performed with 125 µM of tabersonine in the initial 200 µL of YPD/strain/time point.

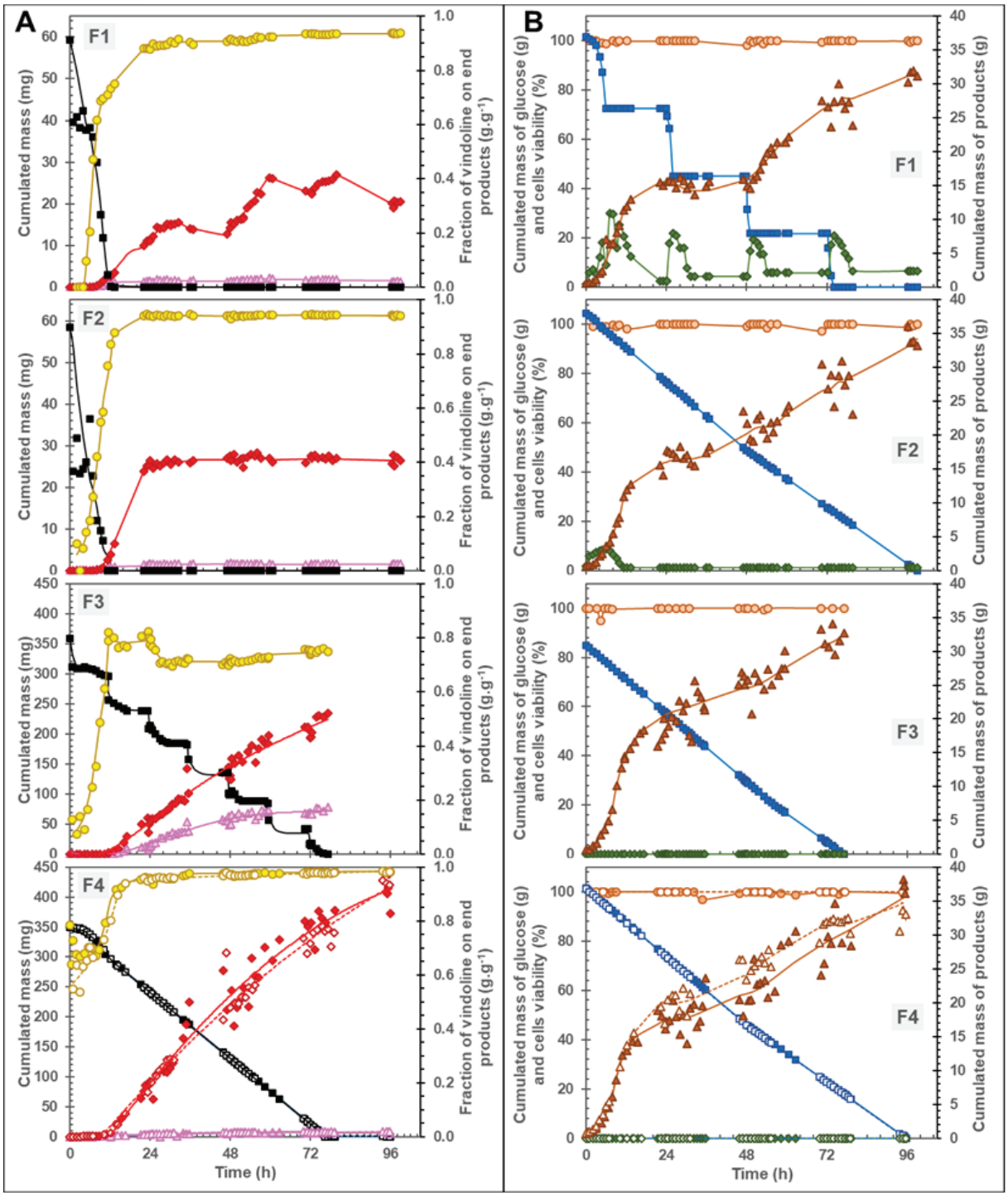


Figure S2: Evolution in bioreactors of (A) cumulated masses of tabersonine, vindoline, vindorosine and the fraction of vindoline on end products, and evolution of (B) cumulated masses of glucose, biomass, ethanol and viability. Consumed tabersonine (dark square), vindoline (red diamond), vindorosine (pink triangle), fraction (yellow circle), consumed glucose (blue square), biomass (brown triangle), ethanol (green diamond) and viability (brown circle). F4 duplicate is in dotted lines and empty symbols. Growth substrates are fed by pulses (F1) or continuously (F2-F4). Tabersonine is fed by one pulse (F1, F2), seven pulses (F3) or equivalent continuous feeding (F4) (Table S4).

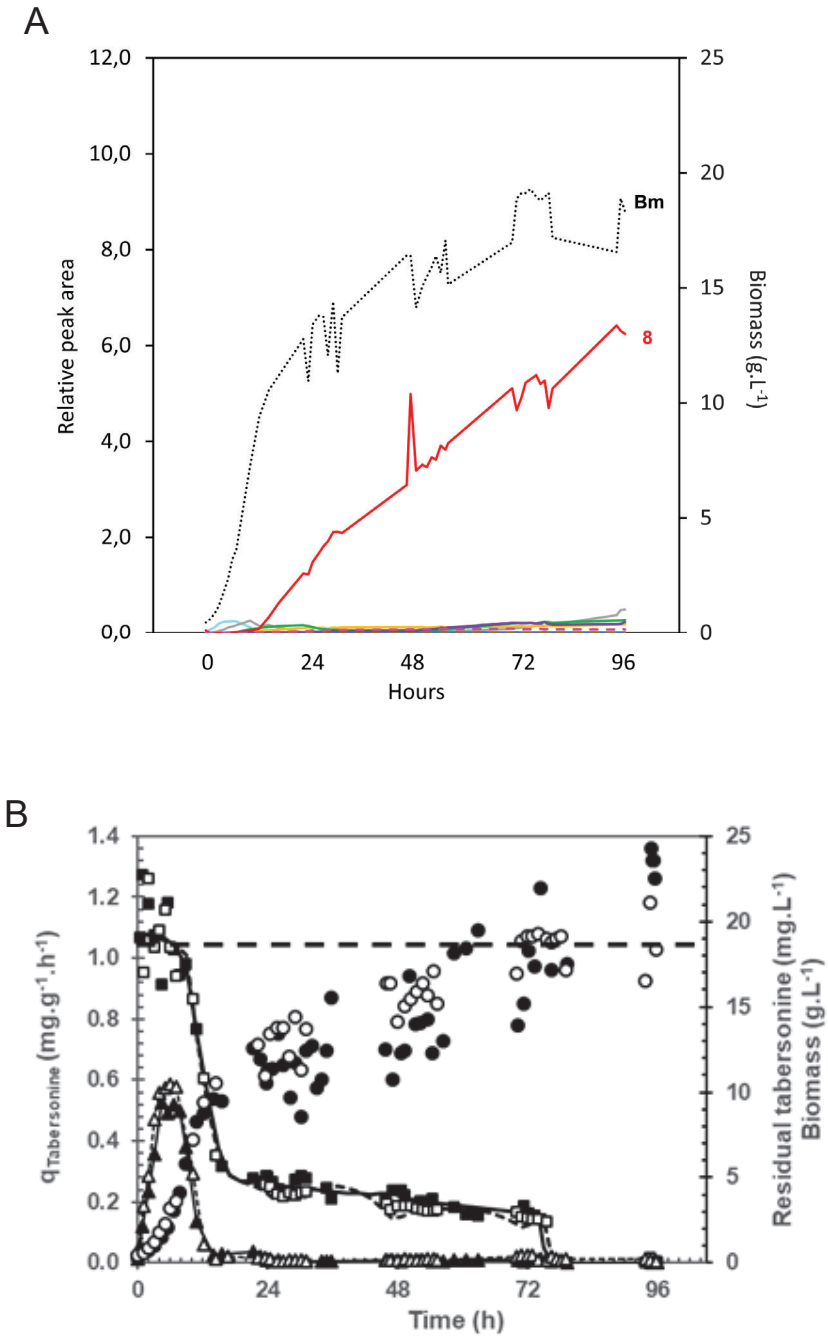


Figure S3: (A) Replicate of the feeding strategy F4 presented in figure 2B. Fresh medium and tabersonine are fed continuously (Table S4). **(B) Evolution of the specific uptake rate of tabersonine in F4 bioreactor.** Specific uptake rate (square), residual tabersonine (triangle) and biomass (circle). The replicate is in dotted line and empty symbols. The horizontal dotted line represents the maximal rate. Tabersonine is fed continuously (Table S4).

Appendix S1: Experimental procedures

Plasmid construction

The vectors used as template donors for the integration of expression cassettes through CRISPR-cas9 method were constructed from a pDONR221 backbone. The Gateway cloning cassette was replaced by a *Not*I restriction site to insert 500 arm sequences upstream and downstream of the different integration loci (Mikkelsen *et al.*, 2012), separated with a *Sal*I restriction site for the cloning of expression cassettes. Promoters and terminators were PCR-amplified from CEN.PK113-7D gDNA (Phusion high fidelity polymerase, Fermentas) using specific primers (**Table S5**) containing overlap sequences, and *Spe*I or *Nhe*I restriction sites for ORF cloning. DNA fragments were purified (Machery-Nagel PCR clean-up kit) and combined (30 ng of promoter and 60 ng of terminator) by overlap PCR using universal primers (Pro_overlap 1 or Pro_overlap2 and Term_T7_SalI or Term_T3_SalI). The obtained promoter/terminator amplicons were purified and combined by a second overlap PCR (40 ng of each promoter/terminator couple) using Term_T7_*Sal*I/Term_T3_*Sal*I primers. The final bidirectional amplicons containing *Sal*I restriction sites at both extremities were digested, purified and ligated into the *Sal*I digested arms-containing backbone vectors creating donor vectors pDPTA111, pDPTA112, pPETA104, pDPTA125, pPEAC115, pDPAC124, pDPAC113, pDPAC122 and pPETA102 for further ORF cloning.

The expression cassettes are composed of two inverted promoter/terminator couples to produce artificial bidirectional promoters, and *Spe*I or *Nhe*I restriction sites were introduced for further ORF cloning. The sequences of tabersonine 16-hydroxylase 1 and 2 (T16H1, T16H2), tabersonine 16-O-methyltransferase (16OMT), tabersonine 3-oxygenase (T3O), tabersonine 3-reductase (T3R), 16-methoxy-2,3-dihydro-3-hydroxytabersonine N-methyltransferase (NMT), deacetoxyvindoline 4-hydroxylase (D4H) and deacetylvindoline O-

acetyltransferase (DAT) from *C. roseus*, as well as NADPH-cytochrome P450 reductase 2 (ATR2) from *Arabidopsis thaliana* were PCR amplified from cDNA, introducing *Spe*I, *Nhe*I or *Xba*I compatible restriction sites for the cloning into expression cassettes. The details of the final template donor vectors are listed in **Table S2**.

Strains

Saccharomyces cerevisiae CEN.PK113-7D (*MATa MAL2-8C, SUC2*) was used as the starting strain (Van Dijken *et al.*, 2000). Single and multiplex CRISPR-Cas9 integration methods were performed to introduce expression cassettes in previously described hotspot yeast loci (Mikkelsen *et al.*, 2012). The CAS9 expression plasmid (pCfB2312) was transformed into *S. cerevisiae* prior the co-transformation with gRNA helper plasmids (pCfB3020, pCfB3041-pCfB3053; (Jessop-Fabre *et al.*, 2016)) and the appropriate NotI-linearized donor DNA vectors, by using the lithium acetate transformation method (Chen *et al.*, 1992). Yeast transformants were selected on YPD medium plates (20 $g.L^{-1}$ of peptone, 10 $g.L^{-1}$ of yeast extract, and 20 $g.L^{-1}$ of glucose) supplemented with G418 (200 $mg.L^{-1}$) and nourseothricin (100 $mg.L^{-1}$), and screened for the effective integration of expression cassette by colony PCR with specific primer couples (**Table S5**).

Small-scale culture conditions

The strains pre-cultured overnight in 5 mL YPD were diluted 20 times in YPD medium, by default prepared with BovMP peptone or, when stated, with different sources of peptone (**Table S3**). Small-scale feedings in final volumes of 200 µL were supplemented with 125 µM of tabersonine and incubated at 28°C, under constant agitation at 200 rpm during 72-120 h, individually for each strain, replicate and time point. Glucose was added every 24h to the cultures to 20 $g.L^{-1}$ (referred to as standard conditions in this study). When stated, from 48 h and every 24 h fresh YP made of BovMP peptone and yeast extract was added (referred to as

YPt_{48h}). When mentioned, YPD medium was buffered to pH 4.0, pH 5.0 or pH 6.0 at t_0 using 50 mM citrate buffers.

Bioreactor culture conditions

The yeast strains were stored in 20% glycerol at -80°C. The cells were first grown on YPD plates (10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ peptone, 20 $g.L^{-1}$ glucose, 15 $g.L^{-1}$ agar) 48 h at 30°C. Three steps of propagation were carried out (5 mL, 15 mL and 150 mL) in YPD from a single colony to inoculate the fed-batch cultures. Each preculture was grown for 10 hours at 30°C, 100 rpm and was used to inoculate the next-step at a 10% ($^{v}/_{v}$) ratio.

Fed-batch cultures were performed in 2 L bioreactors (Global Process Control) with a working volume of 1.5 L, managed with the CBio2 software. The temperature was regulated at 30°C and pH at 6.0 by the addition of 5 M KOH solution. Air flow and stirring rate were adjusted to maintain fully aerobic conditions, i.e. a dissolved oxygen concentration above 20% of saturation. The initial medium composition contained 10 $g.L^{-1}$ yeast extract, 20 $g.L^{-1}$ peptone and, when mentioned, 20 $g.L^{-1}$ glucose. Feeding of nutritive elements was ensured by pumping solutions of glucose (200 g/L) or concentrated YPD (50 $g.L^{-1}$ yeast extract, 100 $g.L^{-1}$ peptone, 200 $g.L^{-1}$ glucose). Feeding of tabersonine was ensured by pulsing a 125 mM solution or pumping a 29 mM one, both prepared in pure ethanol. Four strategies of bioproduction were established, combining pulsed or continuous feeding of glucose, concentrated YPD and tabersonine (**Table S4**). Cells did not receive tabersonine in the F0 reference culture, managed as the F1 for nutrients. F1 was designed to mimic the culture conditions at small scale leading to higher production, by feeding glucose or concentrated YPD by pulses every 24 h. F2 was fed continuously at a steady flow rate with an equivalent amount of nutrients than pulsed in the F1 culture, to avoid any starvation. In F1 and F2, tabersonine was fed with an initial 125 µM pulse when starting the culture. In order to enhance the production of vindoline, six more pulses of 125 µM tabersonine were carried out

every 12 h in F3, based on the F2 nutrients feeding strategy, accordingly. An equivalent amount of tabersonine was continuously fed at a steady flow rate in the F4 culture during the same period of time (from t_0 and until full consumption of tabersonine in F3). F4 was performed in duplicate from two independent colonies.

Sample analysis

In small-scale experiments, supernatants (referred to as extracellular) were obtained by centrifugation (10 min, 7,000g), diluted 20 times in MeOH, vortexed and centrifuged (15 min, 20,000g). In bioreactor experiments, supernatants were obtained by centrifugation of raw samples (4 min, 4,000g) and frozen at −20°C for further analyses.

Determination of tabersonine, vindoline, vindorosine and intermediate metabolites was performed using an UPLC-MS and parameters as previously described (Parage *et al.*, 2016). Retention time and m/z for each compound are listed in **Fig. 1A**. The fragmentations of the main compounds are listed in **Table S1**.

Glucose and ethanol concentrations were analyzed by enzymatic methods with a spectrophotometer and commercial kits (kit 013, kit 026, BioSentec). The accurate determination of glucose, ethanol, glycerol and organic acids from the permeate was performed by high-performance liquid chromatography (Waters), using an Aminex HPX-87H+ column (300 mm * 7.8 mm) set at 50°C, a 5 mM H_2SO_4 as eluent at a flow rate of 0.5 $mL.min^{-1}$ and dual detection by refractometer and UV at 210 nm (Waters).

Determination of biomass

To conduct experiments, biomass was estimated by performing spectrophotometric

measurements at 620 nm with visible spectrophotometer after a calibration against cell dry weight measurements to evaluate yeast growth. For cell dry weight determination, culture medium was harvested and filtered on 0.45 µm pore-size polyamide membranes, which were then dried to a constant weight at 60°C under partial vacuum (200 mmHg, i.e. approximately 26.7 kPa).

Determination of cell viability

Cell coloration with the methylene blue method was used to determine the cell viability (Nielsen et al., 1991). 100 µL of 0.3 mM methyleneblue was mixed with 100 µL of a diluted raw sample and incubated at room temperature for 5 min. A minimum of 200 total cells was counted in at least five different fields of a double-side Malassez counting chamber. The percentage of viable cells was the number of unstained cells (living cells) divided by the total number of cells (stained and unstained). Buds have been taken into account when they seized more than half of the parent-cell.

Off-gas analysis

Off-gas were analyzed by infrared absorption (CO_2) and electrochemical measurement (O_2) with Tandem analyzers (Magellan BioTech).

Table S5: List of primers used in the study

	Name	**Sequence (5'-3')**
Promoter amplification	P_{TEF1}_F	TAGCGAACATGATAATCGACTGGCCACACACCATAGCTTCAAAATG
	P_{TEF1}_R	GAAGTGTCAACAACACTAGTTTTGTAATTAAAACTTAGATTAGATTGC
	P_{PGK1}_F	CCAGTCGATTATCATGTTCGCTAACCCTCATACTATTATCAGGGC

	P_{PGK1}_R	GTTAAAAGCACTGCTAGCTGTTTTATATTTGTTGTAAAAAGTAGATAATTAC
	P_{TDH3}_F	CCAGTCGATTATCATGTTCGCTACATTATCAATACTGCCATTTCAAAGA
	P_{TDH3}_R	GAGCGATTTGTGCTAGCTTTGTTTGTTTATGTGTGTTTATTCG
	P_{ACT1}_F	TAGCGAACATGATAATCGACTGGCTCACCCTAACATATTTTCCAATTAACC
	P_{ACT1}_R	GGTTAGAGCGGATACTAGTTGTTAATTCAGTAAATTTTCGATCTTGG
Terminator amplification	T_{ADH1}_F	CAAAACTAGTGTTGTTGACACTTCTAAATAAGCG
	T_{ADH1}_R	TAATACGACTCACTATAGGGAGCGACCTCATGCTATACC
	T_{CYC1}_F	CTAAACACTAGTATCCGCTCTAACCGAAAAGG
	T_{CYC1}_R	TAATACGACTCACTATAGGGCTTCGAGCGTCCCAAAACCT
	T_{ENO2}_F	ACAGCTAGCAGTGCTTTTAACTAAGAATTATTAGTCTTTTC
	T_{ENO2}_R	GCAATTAACCCTCACTAAAGGGTAGTTATTTTTCAAACTGCAAATTCAAG
	T_{PGI1}_F	CAAAGCTAGCACAAATCGCTCTTAAATATATACCTAAAG
	T_{PGI1}_R	GCAATTAACCCTCACTAAAGGCTTTGAAACTAAATTTGTATATTGTTTGTCAG
Pro/term combination	Pro_overlap 1	CCAGTCGATTATCATGTTCGCTA
	Pro_overlap2	TAGCGAACATGATAATCGACTGG
	Term_T7_SalI	CTGAGAGTCGACTAATACGACTCACTATAGGG
	Term_T3_SalI	CTGAGAGTCGACGCAATTAACCCTCACTAAAGG
ORF amplification	T16H2_SpeI_F	CTGAGAACTAGTTCGGTCATGGAGTTGTATTATTTTTCCACC
	T16H2_SpeI_R	CTGAGAACTAGTCTAATATTTACCTTTGAGAGAAGAAGC
	T16H1_SpeI_F	CTGAGAACTAGTTCGGTCATGGAATTCTATTATTTTCTCTACTTGGC
	T16H1_SpeI_R	CTGAGAACTAGTTCAAGCAGGAGAAGAGGAAG
	16OMT_SpeI_F	CTGAGAACTAGTTCGGTCATGGATGTTCAATCTGAGGAGT
	16OMT_SpeI_R	CTGAGAACTAGTTCAAGGATAAACCTCAATGAGAC
	T3O_SpeI_F	CTGAGAACTAGTTCGGTCATGGAGTTTCATGAATCTTCTCCC
	T3O_SpeI_R	CTGAGAACTAGTTCATGCATAGGACGTAGCGA
	T3R_XbaI_F	CTGAGATCTAGATCGGTCATGGCAATGGCTTCAAAGTC
	T3R_XbaI_R	CTGAGATCTAGATTAGGGTGATTTGAAAGTGTTTCC
	NMT_SpeI_F	CTGAGAACTAGTTCGGTCATGGAAGAGAAGCAGGAGAAG
	NMT_SpeI_R	CTGAGAACTAGTTCATATTGATTTTCGTCCCGTAAC
	D4H_SpeI_F	CTGAGAACTAGTTCGGTCATGCCTAAGTCTTGGCCAAT

	D4H_SpeI_R	CTGAGAACTAGTCTAATTGTTTAACCTGAAAGGAGATAAG
	DAT_XbaI_F	CTGAGATCTAGATCGGTCATGGAGTCAGGAAAAATATCGG
	DAT_XbaI_R	CTGAGATCTAGATTAATTAGAAACAAATTGAAGTAGCTGT
	AtATR2_SpeI_F	CTGAGAACTAGTTCGGTCATGAAAAACATGATGAATTATAAATTAAAAC
	AtATR2_SpeI_R	CTGAGAACTAGTTTACCATACATCTCTAAGATATCTTCC
Screen for effective locus integration	ExpK7_Down_F	CAGGACTAGTGCTGAGGCATTAAT
	XI-1_Down_R	GAAGACCCATGGTTCCAAGGA
	X-4_Down_R	GACGGTACGTTGACCAGAG
	XII-5_Down_R	TGGGCAGCCTTGAGTAAATC
	XI-2_Down_R	GAGACAAGATGGGGCAAGAC
	XII-2_Down_R	GGCCCTGATAAGGTTGTTG
	ExpK7_Up_F	CGAGCTCGCTGAGGACTTAAT
	XI-5_Up_R	TGCGACAGAAGAAAGGGAAG
	XII-4_Up_R	GAACTGACGTCGAAGGCTCT
	XI-3_Up_R	GTGCTTGATTTGCGTCATTC
	X-2_Up_R	TGCGACAGAAGAAAGGGAAG
	XI-2_Up_R	GTTTGTAGTTGGCGGTGGAG
	XII-2_Up_R	CGAAGAAGGCCTGCAATTC

References

Chen, D.C., Yang, B.C., and Kuo, T.T. (1992) One-step transformation of yeast in stationary phase. *Curr Genet* **21**: 83–84.

Colinas, M., Pollier, J., Vaneechoutte, D., Malat, D.G., Schweizer, F., De Milde, L., et al. (2021) Subfunctionalization of paralog transcription factors contributes to regulation of alkaloid pathway branch choice in *Catharanthus roseus*. *Front. Plant Sci.* **12**: 687406.

Van Dijken, J.P., Bauer, J., Brambilla, L., Duboc, P., Francois, J.M., Gancedo, C., et al. (2000) An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains. In *Enzyme and Microbial Technology*. Elsevier Science Inc, pp. 706–714.

Jessop-Fabre, M.M., Jakočiūnas, T., Stovicek, V., Dai, Z., Jensen, M.K., Keasling, J.D., and Borodina, I. (2016) EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. *Biotechnol J* **11**: 1110–1117.

Mikkelsen, M.D., Buron, L.D., Salomonsen, B., Olsen, C.E., Hansen, B.G., Mortensen, U.H., and Halkier, B.A. (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab Eng* **14**: 104–111.

Nielsen, L.K., Smyth, G. K., Greenfield P.F. (1991) Hemacytometer Cell Count Distributions: Implications of Non-Poisson Behavior *Biotechnol. Prog.* 7: 560–563.

Parage, C., Foureau, E., Kellner, F., Burlat, V., Mahroug, S., Lanoue, A., et al. (2016) Class II cytochrome P450 reductase governs the biosynthesis of alkaloids. *Plant Physiol* **172**: 1563–1577.

Schweizer, F., Colinas, M., Pollier, J., Van Moerkercke, A., Vanden Bossche, R., De Clercq, R., and Goossens, A. (2018) An engineered combinatorial module of transcription factors boosts production of monoterpenoid indole alkaloids in *Catharanthus roseus*. *Metab Eng* **48**: 150–162.

IV. Conclusion

The medicinal plants have always been the source of valuable molecules that are continuously used as treatment to different kinds of diseases (Favarin et al., 2013; Erhirhie et al., 2021; Srivastava et al., 2005). However, the low quantity in which these molecules are found in nature is a source of concern in face of the increasing demand from the industries, but not only. Due to the urgent need to protect the environment and slow-down the depletion of resources, alternative methods of production are required according to the Sustainable Development Goals (SDGs) of the United Nations (https://www.un.org/sustainabledevelopment/sustainable-development-goals). Considering that the total chemical synthesis was already discarded as a solution to this problem due to its highly costs and limited technology (Miyazaki et al., 2007), the biosynthetic approach is a sustainable and green alternative to produce plant secondary metabolites. Indeed, the use of microbial factories can offer greater yields and easy extraction processes when compared to the purification of natural products from the host plant in addition to being a sustainable production (Guirimand et al., 2021). The fast improvement of technologies such as the DNA synthesis is currently facilitating the use of heterologous biosynthetic pathways to produce plant natural products by making it faster and cheaper.

However, up to now, the use of the heterologous expression approach for the production of *Catharanthus roseus* alkaloids have been laborious due to the high complexity of MIA biosynthesis. The MIA pathway of *C. roseus* possesses a high degree of compartmentalization in which around 30 enzymatic steps are distributed in 5 different subcellular compartments and in 3 distinct tissues (Courdavault et al., 2014). Fortunately, the MIA biosynthetic pathways of *C. roseus* leading to catharanthine and vindoline have been completely elucidated after 20 years of research, allowing the advancements on the pathway reconstitution of several steps of its MIA synthesis and making possible the accessibility and the sustainable production of anticancer molecules such as vinblastine and vincristine (Caputi et al., 2018).

Heterologous expression involves several steps, starting with the identification of the plant genes that are involved in the biosynthetic pathway of interest, followed by gene validation *in planta*, optimization and culture scaling (figure 17). This benchmark strategy allows the production of plant molecules in microorganisms such as anticancer molecules (Courdavault et al., 2020). In order to summarize each step in the context of this thesis, a brief discussion will be done on the light of the flowchart of figure 17.

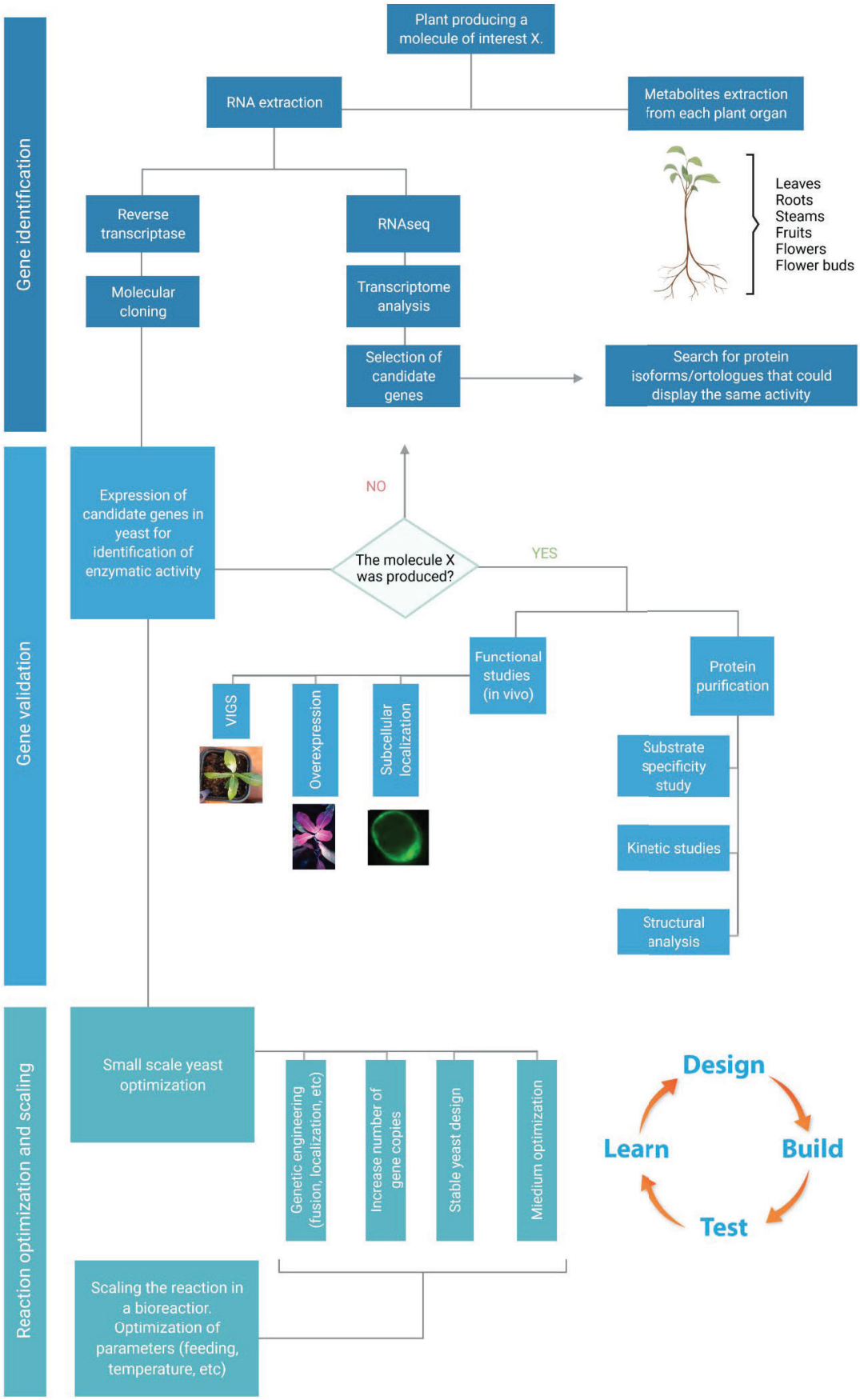


Figure 17: From gene identification to production of natural products in yeast. The flowchart applied in this thesis.

1. Gene validation: methods and considerations

In the part 1, two methods that allow the validation of candidate genes *in vivo* were described. The article 1, described the VIGS technique, in which a candidate gene is silenced on the host plant and its metabolic profile is analysed to inquire the gene's function. The article 2, described the agrobacterium-mediated overexpression, which also allows the validation of a gene's function. Both articles describe essential techniques that give valuable information when studying a new plant gene, especially when the objective is the expression of the gene in a heterologous organism. Indeed, it is impossible to optimize an enzymatic reaction in a heterologous organism without fully understanding the gene's function on its host organism, in this way, these steps should not be left out. The VIGS article is complementary to other publications already describing the VIGS procedure in *C. roseus* and *R. serpentina* by the group in which the thesis was developed in order to update the technique and assemble all the knowledge obtained on the laboratory through several years of development of a simple, effective and transferable VIGS protocol (Carqueijeiro et al., 2015; Corbin et al., 2017). The article 2 introduces an approach for the agrobacteria-mediated overexpression which, despite being a known method and already described for several plant species, was not yet optimized for *C. roseus* with such high efficiency of transformation and applied on MIA pathway elucidation.

The practical application of both methods was put in evidence on the article 3 in which the characterization of a second isoform of the *C. roseus* 16OMT was described. After the biochemical assays and the confirmation of 16OMT activity, the physiological role of each isoform was essential to understand how these proteins behave *in planta* and their importance for the vindoline pathway. The mRNA quantification showed that the second isoform expression was higher than the expression of the first one in every *C. roseus* organ, with the exception of young leaves, where the biggest amount of vindoline is found and their expression was similar. In this way, it was not possible to affirm which isoform would be essential for the MIAs pathway. The VIGS approach could help to answer this question by silencing each isoform followed by the observation of vindoline levels in each condition. However the two isoforms possess a sequence similarity of 86%, which precluded the selection of a region long enough to perform the silencing. After a first try, we observed a cross-silencing and the difficulty of silencing each isoform. Despite the cross-silenced samples, the results showed a great vindoline reduction when both isoforms were silenced, but the initial question was not

answered. Luckly, we had in our hands the very recent approach designed in our laboratory for the overexpression of *C. roseus* genes through agrobacteria-mediated transformation. Since this technique allows the protein production *in planta,* it is highly specific and allows verification of protein expression. Finally, we were able to show that the overexpression of both isoforms led to an increase in the vindoline production and reinforced the importance of both proteins on the MIA pathway.

2. Characterisation of isoformes and orthologues of *C. roseus* MIA genes

In the terms of full comprehension of the gene's behaviour in the plant host, the concept of plant isoforms was introduced, which represents the high complexity of biosynthetic pathways. Taking the example of the already described isoforms of the tabersonine-16-hydroxylase (T16H), the discussion about the existence of several isoforms that may possess different physiological roles was put in place. In the case of T16H, the recently identified isoform (T16H2) was proved to be essential for the vindoline production in small leaves, different from the first one (T16H1). The researchers were able to silence each isoform by the VIGS method and the results showed that the vindoline concentration dropped down when the second isoform was silencing, increasing considerably the production of vindorosine. In addition, the observation of several *C. roseus* cultivars allowed the correlation between the expression profile of the second isoform with the amount of vindoline in leaves (Besseau et al., 2013). This remarkable work shed a light on the evolution of enzymes and how specialized isoforms can emerge within a cell-specific specialized metabolism. This study prompted us to search for the existence of a second isoform of 16OMT, the enzyme catalysing the step following that catalysed by T16H. Interestingly, not only one but two candidate genes called our attention. The first candidate gene 16OMT2 was indeed a second isoform of 16OMT (the previously characterized isoform is 16OMT1) displaying a curious promiscuous activity for flavonoids. This flavonoid promiscuity of 16OMT2 opens the discussion on the evolution process of 16OMT.

The evolution of alkaloids OMTs is complex. Several studies suggests that these OMTs probably evolved from a convergent and recurrent evolution (which can be defined as the repeated evolution of a particular trait, character, or mutation) (Barakat et al., 2011; Lam et al.,

2007; Morris & Facchini, 2019; Schröder et al., 2004). In order to better sustain the discussion about the evolution of the enzymes characterized in this thesis, a second version of the figure 8 from the section "2.4. Evolution of new functionalities in OMTs" was prepared (figure 18). Here, the discussion will be focused on the Apocynaceae OMTs subgroup (shaded in blue). The three proteins elucidated and characterized on this thesis (16OMT2, Vm16OMT and CRO-T104094) are highlighted in bold blue.

As expected, all the three proteins are allocated within the same subgroup, in which alkaloids and flavonoids OMTs from *C. roseus* are present. Such as 16OMT1, 16OMT2 shows a high specificity for the alkaloid 16-hydroxytabersonine. However the particular promiscuity of 16OMT2 for flavonoids suggests that this protein may have evolved from a gene duplication of an ancestral flavonoid OMT gene. The lack of flavonoid OMT activity of 16OMT1 suggests that this gene appeared afterwards, being the result of a later gene duplication and neofunctionalization of 16OMT2.

The figure 18 also highlights the phylogenetic proximity between 16OMT1, 16OMT2 and Vm16OMT. If 16OMT1 raised from a continuous neofunctionalization of 16OMT2, it would be reasonable to hypothesize that the ancestral gene of 16OMT1, 16OMT2 and Vm16OMT was a protein capable of methylating both flavonoids and alkaloids. It would be interesting to check if Vm16OMT still accepts flavonoids as substrates to reinforce this hypothesis. Considering the presence of several flavonoids OMTs in this subgroup, this could be a strong hypothesis.

The chances for gene duplicated sequences to evolve to new functions are small compared with the probability that the copies become disabled by unfavourable mutations (Rodriguez-Trelles et al., 2003). Despite possessing almost 70% amino acid identity with 16OMT2, the second candidate gene identified to be a 16OMT isoform (CRO-T104094) did not display methyltransferase activity with any of the substrate tested. According to its location in the tree (figure 18), this gene also resulted from the duplication and mutation of the same gene ancestor that gave rise to this subgroup.

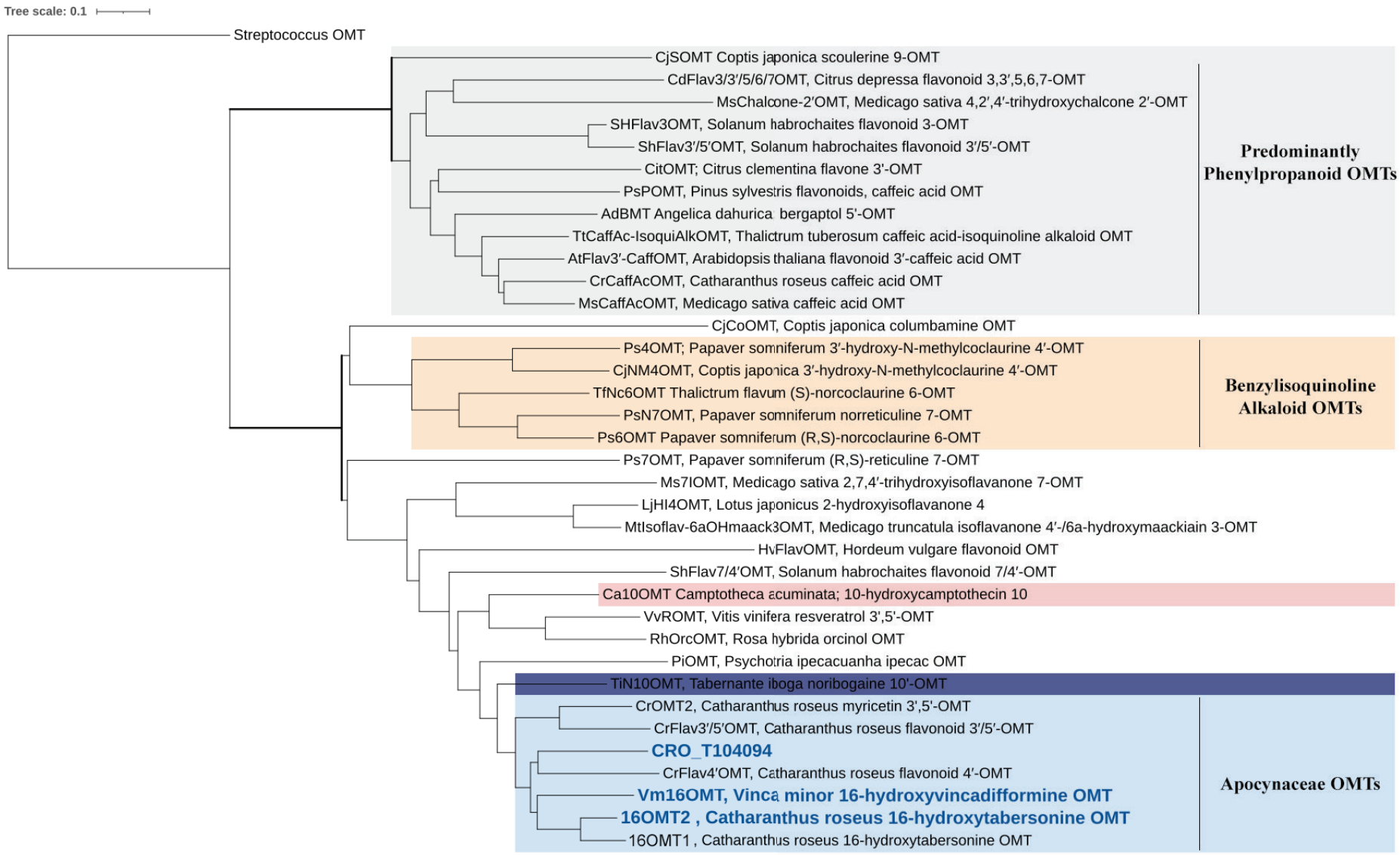


Figure 18: Updated Phylogenetic tree with the genes studied in this thesis.

Proteins with no reported activity *in vitro* may also display other kinds of function. For example, the recently reported shSGD in *C. roseus* is a shorter version of the known protein SGD (strictosidine-β-glucosidase) (Carqueijeiro et al., 2021). The shSGD does not possess activity against any substrate tested up to date, however, when this enzyme is co-expressed with SGD, it disrupts SGD multimers, reducing its activity both *in vitro* and *in planta*. The authors classified this enzyme as a "zombie protein", which can be defined as a protein with no catalytic reaction associated but may interact with other proteins in order to regulate its activity. In the article 3, experiments to search for *in vivo* activity of CRO-T104094 were performed *via* agrobacterium-mediated overexpression, but no activity was observed for alkaloid biosynthesis. Further experiments should be performed in the future in order to investigate a possible zombie activity for CRO-T104094.

The presence of CrFlavo4'OMT in this subgroup could provide some information that could help on future investigations of CRO-T104094. CrFlavo4'OMT is a particular flavonoid OMT described by Schröder et al. (2004). The only substrate accepted by this enzyme *in vitro*, up to now, is the flavanone 3'-*O*-methyl-eriodictyol (homoeriodictyol) and its role in *C. roseus* metabolism is still unknown. Noting that 3',4'-O-methylated compounds have not been described in *C. roseus* (Schröder et al., 2004). Surprisingly, the CrFlavo4'OMT methylates the B-ring of homoeriodictyol, a mechanism that was not yet reported for other plant OMTs. These distinct features of CrFlavo4'OMT and its phylogenetic proximity to CRO-T104094 could suggest that CRO-T104094 accepts a substrate not yet described in *C. roseus*. Further experimentation could be performed, however the lack of solubility of the enzyme would definitely be a source of constraint.

3. Production of vindoline by yeast metabolic engineering

In the part 3, two articles focused on the metabolic engineering of the yeast producing vindoline are presented. The article 5 presents practical strategies to optimize a biosynthetic pathway in yeast, focusing on the intermediates of the vindoline pathway. This small-scale optimization is mandatory to assure a maximum conversion of each intermediate and the bottle-neck solving. The article 6 presents the final yeast vindoline producer and its optimization in a bioreactor, which was achieved by following each step listed on the figure 17. This happy end

story allowed us to put in place this benchmark procedure that can be transferred to other plants and other molecules, in order to advance the production of molecules using yeast cell factories.

Overall, the identification of Vm16OMT (in article 4) it is not only important for the complete elucidation of the *V. minor* MIA pathway, but also, the isoforms and orthologues identification is of great interest of metabolic engineering, due to the possible use of these proteins in combination for metabolic flux optimization as presented on the article 5. In this article, several molecular strategies are used in order to solve a bottleneck on the reaction flux of a biosynthetic pathway. In the case of vindoline production, the methylation of 16-hydroxytabersonine was proven to be inefficient, which led to an accumulation of this last intermediate and a severe reduction in intermediates production. The use of Vm16OMT in the place of 16OMT1 was performed, despite no improvement on the conversion was observed. This result was not surprising considering the fact that *V. minor* does not produce vindoline and the promiscuity of this enzyme for 16-hydroxytabersonine is a feature shared to the close relation of both orthologues. However, this does not exclude the use of orthologues for metabolic engineering. A successful example of orthologue use can be found in the recent article of Christina Smolke's group, which describes the biosynthesis of two valuable medicinal tropane alkaloids in yeast at laboratory scale by no less than 34 genome editions. The strategy involved orthologue identification and testing of several enzymes across species to find the most active ones (Srinivasan & Smolke, 2019). This example shows the importance of exploring the natural sequence diversity for pathway engineering strategies.

In addition to orthologue use, another molecular strategy was put in place on article 5. The increased number of gene copies is a known strategy that can rapidly improve the metabolic flux, however some concerns need to be taken in consideration. In the case of vindoline pathway, we observed the presence of several promiscuous enzymes so the addition of one extra copy needs to be done in order to promote the production of the desired metabolite. We added an extra copy of T16H2 to favour the vindoline production instead of the by-product vindorosine (Figure 19). After the addition of a second copy of T16H2, a huge accumulation of its product (16-hydroxytabersonine) was observed. After several tests, we decided to equally add an extra copy of 16OMT1, which increased considerably the 16-methoxytabersonine production. One may notice a slight accumulation of 16-methoxytabersonine when a second copy of 16OMT was added, but no added copy of the next enzyme of the pathway (T3O) was considered. Figure 19 highlights the role of T3O on vindoline and vindorosine pathways. This enzyme has a high specificity for both tabersonine and 16-methoxytabersonine and an extra

copy of this gene would increase the competitiveness toward the vindorosine pathway, which was exactly what we succeeded to minimize by adding an extra copy of T16H.

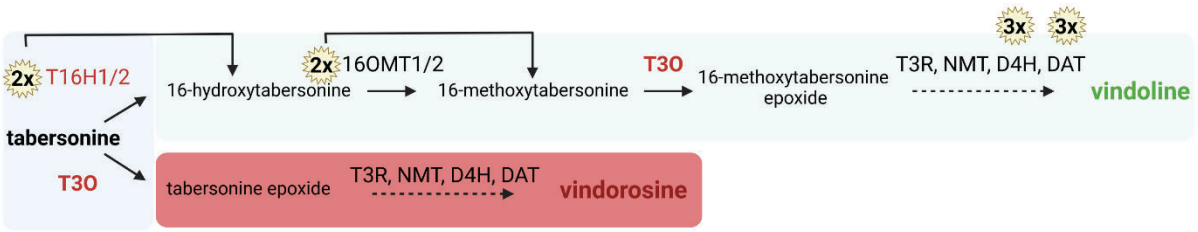


Figure 19: Optimization of the pathway for vindoline production.

Besides the optimization of intermediates production by the identification of more active enzymes and increasing the number of gene copies, media optimization and culture conditions needs to be implemented in order to increase vindoline production. The article 6 shows how these parameters can drastically improve the vindoline production especially in bioreactors. On the first bioreactor assays of the article 6, the optimize parameters from article 5 were kept, meaning the 2 copies of T16H and 2 copies of 16OMT. However, due to the changes on the technique used for the genome edition (from homologous recombination to CRISPR-Cas9), we decided to make use of the two gene isoforms, as the second copy, to avoid possible recombination. Hence, T16H1 and T16H2 were used, as well as, 16OMT1 and the recently described on this thesis, 16OMT2. The use of two copies of the same isoform or the combination between them does not change the bioconversion of tabersonine to vindoline. Thus, the choice of using both isoforms of each gene was made solely based on the creation of a highly stable strain.

Notwithstanding the high optimization of the vindoline intermediate production of the article 5, further molecular modifications on the yeast strains were required. When analysing the vindoline producer yeast strain (article 6), we observed an accumulation of desacetoxyvindoline that prompted us to increase the D4H and DAT copy number (figure 19). Despite being also promiscuous enzymes that take part on the vindorosine pathway, the increase of the number of copies of these genes were not a problem since we have considerably reduced the production of the first metabolite of vindorosine pathway, tabersonine epoxide. In this way, we could increase the copy number of both genes until we saw no further improvement on the conversion, up to 3 final copies of each gene. The limits of the approach of gene copies increase relies on the cofactor availability related to the endogenous yeast

metabolism. Both D4H and DAT are enzymes that rely on cofactor availability, 2-oxoglutarate and acetyl-CoA respectively. It would be possible to proceed to further enhancements on yeast metabolism by adding extra-copies of endogenous yeast genes for co-factor production. For example, in the article 5, we tested the possibility of increasing the 16OMT1 activity by adding an extra copy of the gene *SAM2* responsible for the synthesis of the cofactor S-adenosyl methionine together with the addition of methionine on the yeast media, but no improvement on the methylation yield was observed (data not published).

Lastly, the media optimization, pulse feeding, oxygenation and nutrition of yeast in bioreactors were essential for the beginning of the scale up of our vindoline bioproduction. Plenty of work is still to be done in order to reach the industrial scale, but such a high synthesis of vindoline opens the door to a diversification of the production of alkaloids though precursor-directed biosynthesis.

V. Perspectives

The high-scale heterologous production of plant molecules encounters an extensive range of obstacles due to the high complexity of plant metabolic pathways as shown throughout this thesis. We presented a multidisciplinary study of the MIA pathways of *C. roseus*, which was extended to other Apocynaceae plants: *R. serpentina* and *V. minor*. Several specific features of *C. roseus* pathway were presented and discussed such as multi-enzyme pathway subcellular compartmentalization, enzyme promiscuity, protein isoforms, enzyme activity and competing native metabolism (for example the pathways leading to vindoline and vindorosine). The full understanding of these mechanisms allowed us to explore and optimize the vindoline's pathway on the heterologous organism *S. cerevisiae*. With the use of transcriptomics, molecular biology and bioprocess tools, we were able to present a successful vindoline production at the laboratory bioreactor scale.

Many *C. roseus* biosynthetic genes are now fully characterized, which allowed us to use its transcriptomics data to identify genes in other Apocynaceae species such as *V. minor* by using guilt-by-association studies (genes with similar expression patterns are likely to be functionally related) together with the traditional homology-based approach. The continuous gene identification and characterization drives the future supervised machine learning (SML)-based approaches to predict gene function (Dugé de Bernonville et al., 2020; Kim et al., 2020). These approaches are considered promising, however further optimization is required to handle the identification of genes that do not follow a certain pattern of homogeneity (co-expression, physical clustering, or sequence homology) (Dugé de Bernonville et al., 2020). In this case, more conventional steps are required to elucidate the gene's function such as native protein purification, or *in planta* approaches such as the VIGS technique, which was described and optimized for two species on this thesis: *C. roseus* and *R. serpentina*. Complementary to the VIGS analysis, the hereby described agrobacterium-mediated overexpression technique is also a high-output alternative for gene validation in *C. roseus.* This elucidation and validation of biosynthetic genes remains an essential prerequisite for the alternative production of valuable pharmaceutical compounds in microbial cell factories and hopefully, the technology advancements will make it easier to validate candidate genes *in planta* with new fast and reliable strategies.

In addition to metabolic pathway elucidation, the development of synthetic biology tools and the studies in metabolic engineering, including fine-tuning of gene expression, orthologue use, cofactor and substrate supply, as well as exploiting subcellular localization, enabled the vindoline pathway reconstitution and its optimization in yeast. Surely, several other

parameters could be further optimized for vindoline production in yeast. For instance, the optimization of native metabolism is frequently employed to rewire metabolic fluxes to avoid co-factor and/or substrate competition (Luo et al., 2019; Dusséaux et al., 2020; Pyne et al., 2020; Srinivasan and Smolke, 2020). This approach would promote steps in which NADPH-dependent reductases and dehydrogenases are involved. In the case of vindoline production, we observed a lack of co-factor availability when additional copies of DAT were added, confirming that the enzymatic step catalysed by DAT would benefit from an increase of CoA availability. For this matter, the DAT re-localization to the peroxisome appears as a promising strategy to increase it yields considering the high beta-oxidation rates inside this organelle (Dusséaux et al., 2020; Kulagina et al., 2021). In parallel, other strategies rely on the enlargement of the endoplasmic reticulum size by a mutation on the PAH1 gene (phosphatidic acid phosphatase) (Carman & Han, 2019) in order to benefit P450s, such as T16H.

All these new strategies relate to new gene insertions in the host genome. Hence, an expansion of high-quality transcriptomic data and of metabolic engineering tools are required for such extensive genetic modifications. Here, our final yeast strain was built with the CRISPR/Cas9 system, a technology which is currently advancing in large steps. Our final yeast strain Hep3 for vindoline production contains 14 gene insertions in 7 different integration locus. More recently, 8 new integration locus were described in *Saccharomyces* which will facilitate the insertion of many other genes that are needed to proceed with the optimization of this pathway (Babaei et al., 2021). These advances also pave the way for the *de novo* production of MIAs which has been a technological challenge in recent years due to the high complexity and number of enzymatic steps involved on the pathway. Advances have been made in this area through proofs of concept showing the biosynthesis of strictosidine in the heterologous plant *Nicotiana benthamiana* (Miettinen et al., 2014) and the de novo production of strictosidine by *S. cerevisiae* (Brown et al., 2015).

The optimization of yeast culture media ingredients is also of great importance regarding the industrialization of this process as they can account for a significant part of the process costs. At the laboratory bioreactor scale, YPD (yeast extract peptone dextrose) media is used as it is one of the richest media available for yeast. This media is costly and not optimized for the needs of the vindoline producing strain. One way to decrease feedstock cost could be to find and adapt cheap carbon sources from agricultural by-products, which could also contribute to environmental sustainability. Recently, *S. cerevisiae* and other different

yeasts were cultivated in a media composed of enzymatically saccharified sulfite-pulped spruce wood and hydrolysates of by-products from chicken (Lapeña et al., 2020).

To conclude, the advances in biosynthetic pathway elucidation and metabolic engineering are not only limited to producing natural metabolites. Indeed, the identification and characterization of new promiscuous enzymes could serve as a tool for non-natural molecules tailoring, as shown by Carqueijeiro et al. (2018), by using the TAT enzyme for the production of 16-methoxy-19-acetyltabersonine.

Lastly, nature is extremely rich and new therapeutic molecules are about to be discovered in many other plants, animals and marine organisms. Surely, all these multidisciplinary advances pave the way for achieving the production of natural and non-natural molecules that could be used in new therapeutic solutions.

Perspectives en français

Comme le montre cette thèse, la production hétérologue à grande échelle de molécules végétales se heurte à de nombreux obstacles en raison de la grande complexité des voies métaboliques des plantes. Nous avons présenté une étude multidisciplinaire de la voie des AIM de *C. roseus*, qui a été étendue à d'autres plantes de la famille des Apocynaceae : *R. serpentina* et *V. minor*. Plusieurs caractéristiques spécifiques de la voie des AIM de *C. roseus* ont été présentées et discutées dont la compartimentation subcellulaire de la voie, la compétition entre chaînes métaboliques (par exemple celles menant à la vindorosine et à la vindoline) et les caractéristiques liées aux enzymes (promiscuité, isoformes, activité). La compréhension approfondie de ces mécanismes nous a permis d'explorer et d'optimiser la voie de la vindoline dans la levure *S. cerevisiae*. L'utilisation de la transcriptomique, d'outils de biologie moléculaire et la maîtrise de bioprocédés, nous a conduit à produire avec succès de la vindoline à l'échelle du bioréacteur de laboratoire.

De nombreux gènes biosynthétiques de *C. roseus* sont dorénavant entièrement caractérisés. Ceci nous a permis d'utiliser les données transcriptomiques de la plante pour identifier des gènes d'autres espèces d'Apocynacées telles que *V. minor* en réalisant des études basées sur le principe de « guilt by association » (les gènes ayant des profils d'expression similaires sont susceptibles d'être fonctionnellement liés) et en utilisant des approches plus

traditionnelles basées sur l'homologie de séquence. L'identification et la caractérisation continue des gènes sont à l'origine de futures approches basées sur l'apprentissage supervisé (SML : Supervised Machine Learning) pour prédire la fonction des gènes (Dugé de Bernonville et al., 2020; Kim et al., 2020). Ces approches sont considérées comme prometteuses, mais nécessitent davantage d'optimisation pour identifier des gènes qui ne suivent pas un certain schéma d'homogénéité (co-expression, gènes voisins sur le génome formant un cluster ou homologies de séquence) (Dugé de Bernonville et al., 2020). Dans ce cas, des étapes plus conventionnelles sont nécessaires pour élucider la fonction de ces gènes comme la purification de protéines natives ou encore des approches *in planta* comme la technique VIGS, qui a été décrite et optimisée pour deux espèces dans cette thèse : *C. roseus* et *R. serpentina*. En complément de l'analyse VIGS, la technique de surexpression induite par les agrobactéries est également une alternative pour la validation des gènes chez *C. roseus*. Cette élucidation et validation des gènes biosynthétiques restent une condition préalable essentielle à la production alternative de composés pharmaceutiques à haute valeur ajoutée dans les usines cellulaires microbiennes. Par ailleurs, on peut espérer que les progrès technologiques faciliteront la validation des gènes candidats *in planta* en ayant recours à de nouvelles stratégies rapides et fiables.

Outre l'élucidation des voies de biosynthèse, le développement d'outils de biologie synthétique et les études menées en ingénierie métabolique, comprenant le réglage fin de l'expression des gènes, l'utilisation d'orthologues, l'approvisionnement en substrats et cofacteurs ainsi que l'exploitation de la localisation subcellulaire, ont permis de reconstituer et d'optimiser la voie de la vindoline dans la levure. Certes, plusieurs autres paramètres pourraient être davantage optimisés pour la production de vindoline dans la levure. Par exemple, l'optimisation du métabolisme natif est fréquemment utilisée pour réorienter les flux métaboliques en modifiant les voies métaboliques pour éviter la compétition entre co-facteurs et/ou substrats (Luo et al., 2019 ; Dusséaux et al., 2020 ; Pyne et al., 2020 ; Srinivasan et Smolke, 2020). Cette approche favoriserait des étapes dans lesquelles les réductases et déshydrogénases dépendantes du NADPH sont impliquées. Dans le cas de la production de vindoline, nous avons observé un manque de disponibilité des cofacteurs lors de l'ajout de copies supplémentaires de DAT, confirmant que l'étape enzymatique catalysée par DAT bénéficierait d'une augmentation de la disponibilité en CoA. A ce titre, la relocalisation de DAT vers le peroxysome apparaît comme une stratégie prometteuse pour augmenter ses rendements compte tenu des taux élevés de bêta-oxydation au sein de cet organite (Dusséaux et al., 2020 ;

Grewal et al., 2021, Kulagina et al. , 2021). En parallèle, d'autres stratégies reposent sur l'agrandissement de la taille du réticulum endoplasmique par une mutation du gène PAH1 (phosphatidate phosphatase) (Carman & Han, 2019) afin de bénéficier aux P450, comme la T16H.

Toutes ces nouvelles stratégies concernent de nouvelles insertions de gènes dans le génome de l'hôte. Par conséquent, une expansion des données transcriptomiques de haute qualité et des outils d'ingénierie métabolique sont nécessaires pour ces importantes modifications génétiques (Lee et al., 2012; Zhang et al., 2020). Dans le cadre de cette thèse, notre souche de levure finale a été élaborée avec le système CRISPR/Cas9, une technologie qui progresse actuellement à grands pas. La souche de levure Hep3 pour la production de vindoline contient 14 insertions de gènes dans 7 locus d'intégration différents. Plus récemment, 8 nouveaux locus d'intégration ont été décrits chez Saccharomyces qui faciliteront l'insertion de nombreux autres gènes nécessaires pour procéder à l'optimisation de cette voie de biosynthèse (Babaei et al., 2021). Ces avancées ouvrent également la voie à la production *de novo* d'AIM qui a été un défi technologique ces dernières années en raison de la grande complexité et du nombre d'étapes enzymatiques impliquées dans la voie. Des avancées ont été réalisées dans ce domaine à travers des preuves de concept montrant la biosynthèse de strictosidine dans la plante hétérologue *Nicotiana benthamiana* (Miettinen et al., 2014) et la production *de novo* de strictosidine par *S. cerevisiae* (Brown et al., 2015).

L'optimisation des ingrédients des milieux de culture des levures est également d'une grande importance pour l'industrialisation de ce procédé car ils peuvent représenter une part importante des coûts du procédé. A l'échelle du bioréacteur de laboratoire, le milieu YPD (Yeast Extract Peptone Dextrose) est utilisé car il s'agit de l'un des milieux les plus riches disponible pour la levure. Ce milieu est coûteux et non optimisé pour les besoins de la souche productrice de vindoline. Une façon de réduire le coût des matières premières pourrait être de trouver et d'adapter des sources de carbone bon marché à partir de sous-produits agricoles, ce qui pourrait également contribuer à la viabilité environnementale. Récemment, *S. cerevisiae* et d'autres levures ont été cultivées dans un milieu composé de pulpe de bois d'épicéa enzymatiquement saccharifiée et d'hydrolysats de sous-produits de poulet (Lapeña et al., 2020).

Pour conclure, les progrès dans l'élucidation des voies de biosynthèse et en ingénierie métabolique ne se limitent pas seulement à produire des métabolites naturels. En effet, l'identification et la caractérisation de nouvelles enzymes de promiscuité pourraient servir

d'outil pour l'élaboration de molécules non naturelles, comme l'ont montré Carqueijeiro et al. (2018) en utilisant l'enzyme TAT pour la production de 16-méthoxy-19-acétyltabersonine.

Enfin, la nature est extrêmement riche et de nouvelles molécules thérapeutiques sont sur le point d'être découvertes dans de nombreuses autres plantes, animaux et organismes marins. Assurément, toutes ces avancées multidisciplinaires ouvrent la voie à la réalisation de la production de molécules naturelles et non naturelles qui pourraient être utilisées dans de nouvelles solutions thérapeutiques.

VI. References

A

Achan, J., Talisuna, A. O., Erhart, A., Yeka, A., Tibenderana, J. K., Baliraine, F. N., Rosenthal, P. J., & D'Alessandro, U. (2011). Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malaria Journal*, *10*(1), 144. https://doi.org/10.1186/1475-2875-10-144

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., & Stephanopoulos, G. (2010). Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science*, *330*(6000), 70–74. https://doi.org/10.1126/science.1191652

Akashi, T., Aoki, T., & Ayabe, S. (2005). Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. *Plant Physiology*, *137*(3), 882–891. https://doi.org/10.1104/pp.104.056747

Alam, M. M., Naeem, M., Khan, M. M. A., & Uddin, M. (2017). Vincristine and vinblastine anticancer *Catharanthus* alkaloids: pharmacological applications and strategies for yield improvement. In *Catharanthus roseus* (pp. 277–307). Springer International Publishing. https://doi.org/10.1007/978-3-319-51620-2_11

Almagro, L., Fernández-Pérez, F., & Pedreño, M. (2015). Indole alkaloids from *Catharanthus roseus*: bioproduction and their effect on human health. *Molecules*, *20*(2), 2973–3000. https://doi.org/10.3390/molecules20022973

Ashihara, H., Sano, H., & Crozier, A. (2008). Caffeine and related purine alkaloids: Biosynthesis, catabolism, function and genetic engineering. *Phytochemistry*, *69*(4), 841–856. https://doi.org/10.1016/j.phytochem.2007.10.029

B

Babaei, M., Sartori, L., Karpukhin, A., Abashkin, D., Matrosova, E., & Borodina, I. (2021). Expansion of EasyClone-MarkerFree toolkit for *saccharomyces cerevisiae* genome with new integration sites. *FEMS Yeast Research*, *21*(4). https://doi.org/10.1093/femsyr/foab027

Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., dePamphilis, C., Albert, V. A., Aono, N., Aoyama, T., Ambrose, B. A., Ashton, N. W., Axtell, M. J., Barker, E., Barker, M. S., Bennetzen, J. L., Bonawitz, N. D., Chapple, C., Cheng, C., Correa, L. G. G., Grigoriev, I. V. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science*, *332*(6032), 960 LP – 963. https://doi.org/10.1126/science.1203810

Barakat, A., Choi, A., Yassin, N. B. M., Park, J. S., Sun, Z., & Carlson, J. E. (2011). Comparative genomics and evolutionary analyses of the O-methyltransferase gene family in *Populus*. *Gene*, *479*(1–2), 37–46. https://doi.org/10.1016/j.gene.2011.02.008

Barreiro, E. J., Kümmerle, A. E., & Fraga, C. A. M. (2011). The methylation effect in medicinal chemistry. *Chemical Reviews*, *111*(9), 5215–5246. https://doi.org/10.1021/cr200060g

Besseau, S., Kellner, F., Lanoue, A., Thamm, A. M. K., Salim, V., Schneider, B., Geu-Flores, F., Höfer, R., Guirimand, G., Guihur, A., Oudin, A., Glevarec, G., Foureau, E., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B., Werck-Reichhart, D., Burlat, V., Courdavault, V. (2013). A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant Physiology*, *163*(4), 1792–1803. https://doi.org/10.1104/pp.113.222828

Bhambhani, S., Kondhare, K. R., & Giri, A. P. (2021). Diversity in chemical structures and biological properties of plant alkaloids. *Molecules*, *26*(11), 3374. https://doi.org/10.3390/molecules26113374

Bojić, M., Maleš, Ž., Antolić, A., Babić, I., & Tomičić, M. (2019). Antithrombotic activity of flavonoids and polyphenols rich plant species. *Acta Pharmaceutica*, *69*(4), 483–495. https://doi.org/10.2478/acph-2019-0050

Bornscheuer, U. T., & Kazlauskas, R. J. (2004). Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angewandte Chemie International Edition*, *43*(45), 6032–6040. https://doi.org/10.1002/anie.200460416

Bourgaud, F., Gravot, A., Milesi, S., & Gontier, E. (2001). Production of plant secondary metabolites: a historical perspective. *Plant Science*, *161*(5), 839–851. https://doi.org/10.1016/S0168-9452(01)00490-3

Brown, D. E., Rashotte, A. M., Murphy, A. S., Normanly, J., Tague, B. W., Peer, W. A., Taiz, L., & Muday, G. K. (2001). Flavonoids act as negative regulators of auxin transport in vivo in *Arabidopsis*. *Plant Physiology*, *126*(2), 524–535. https://doi.org/10.1104/pp.126.2.524

Brown, S., Clastre, M., Courdavault, V., & O'Connor, S. E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences*, *112*(11), 3205–3210. https://doi.org/10.1073/pnas.1423555112

Buer, C. S., & Muday, G. K. (2004). The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of *Arabidopsis* roots to gravity and light. *The Plant Cell*, *16*(5), 1191–1205. https://doi.org/10.1105/tpc.020313

Buer, C. S., Sukumar, P., & Muday, G. K. (2006). Ethylene modulates flavonoid accumulation and gravitropic responses in roots of *Arabidopsis*. *Plant Physiology*, *140*(4), 1384–1396. https://doi.org/10.1104/pp.105.075671

Byeon, Y., Lee, H., Lee, K., & Back, K. (2014). Caffeic acid O-methyltransferase is involved in the synthesis of melatonin by methylating *N*-acetylserotonin in *Arabidopsis*. *J Pineal Res*, *57*(2), 219–227. https://doi.org/10.1111/jpi.12160

Byeon, Y., Lee, H.-J., Lee, H., & Back, K. (2016). Cloning and functional characterization of the *Arabidopsis* N-acetylserotonin O-methyltransferase responsible for melatonin synthesis. *J Pineal Res*, *60*(1), 65–73. https://doi.org/10.1111/jpi.12289

C

Cacace, S., Schröder, G., Wehinger, E., Strack, D., Schmidt, J., & Schröder, J. (2003). A flavonol O-methyltransferase from *Catharanthus roseus* performing two sequential

methylations. *Phytochemistry*, *62*(2), 127–137. https://doi.org/https://doi.org/10.1016/S0031-9422(02)00483-1

Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T. D., Dang, T. T. T., Carqueijeiro, I. S. T., Koudounas, K., de Bernonville, T. D., Ameyaw, B., Jones, D. M., Vieira, I. J. C., Courdavault, V., & O'Connor, S. E. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science*, *360(6394)*, 1235-1239. https://doi.org/10.1126/science.aat4100

Carlin, M. G., Dean, J. R., & Ames, J. M. (2020). Opium alkaloids in harvested and thermally processed poppy seeds. *Frontiers in Chemistry*, *8:*737. https://doi.org/10.3389/fchem.2020.00737

Carman, G.M. & Han, G.S. (2019). Fat-regulating phosphatidic acid phosphatase: a review of its roles and regulation in lipid homeostasis. *Journal of Lipid Research*, *60*, 2–6. https://doi.org/10.1194/jlr.S087452.

Carqueijeiro, I., Dugé de Bernonville, T., Lanoue, A., Dang, T.-T., Teijaro, C. N., Paetz, C., Billet, K., Mosquera, A., Oudin, A., Besseau, S., Papon, N., Glévarec, G., Atehortùa, L., Clastre, M., Giglioli-Guivarc'h, N., Schneider, B., St-Pierre, B., Andrade, R. B., O'Connor, S. E., & Courdavault, V. (2018). A BAHD acyltransferase catalyzing 19-O-acetylation of tabersonine derivatives in roots of *Catharanthus roseus* enables combinatorial synthesis of monoterpene indole alkaloids. *The Plant Journal*, *94*(3), 469–484. https://doi.org/10.1111/tpj.13868

Carqueijeiro, I., Guimarães, A. L., Bettencourt, S., Martínez-Cortés, T., Guedes, J. G., Gardner, R., Lopes, T., Andrade, C., Bispo, C., Martins, N. P., Andrade, P., Valentão, P., Valente, I. M., Rodrigues, J. A., Duarte, P., & Sottomayor, M. (2016). Isolation of cells specialized in anticancer alkaloid metabolism by fluorescence-activated cell sorting. *Plant Physiology*, *171*(4), 2371–2378. https://doi.org/10.1104/pp.16.01028

Carqueijeiro, I., Koudounas, K., Dugé de Bernonville, T., Sepúlveda, L. J., Mosquera, A., Bomzan, D. P., Oudin, A., Lanoue, A., Besseau, S., Lemos Cruz, P., Kulagina, N., Stander, E. A., Eymieux, S., Burlaud-Gaillard, J., Blanchard, E., Clastre, M., Atehortùa, L., St-Pierre, B., Giglioli-Guivarc'h, N., Courdavault, V. (2021). Alternative splicing creates a pseudo-strictosidine β-d-glucosidase modulating alkaloid synthesis in *Catharanthus roseus*. *Plant Physiology*, *185*(3), 836–856. https://doi.org/10.1093/plphys/kiaa075

Carqueijeiro, I., Masini, E., Foureau, E., Sepúlveda, L. J., Marais, E., Lanoue, A., Besseau, S., Papon, N., Clastre, M., Dugé de Bernonville, T., Glévarec, G., Atehortùa, L., Oudin, A., & Courdavault, V. (2015). Virus-induced gene silencing in *Catharanthus roseus* by biolistic inoculation of tobacco rattle virus vectors. *Plant Biology*, *17*(6), 1242–1246. https://doi.org/10.1111/plb.12380

Copley, S. D. (2015). An evolutionary biochemist's perspective on promiscuity. *Trends in Biochemical Sciences*, *40*(2), 72–78. https://doi.org/10.1016/j.tibs.2014.12.004

Corbin, C., Lafontaine, F., Sepúlveda, L. J., Carqueijeiro, I., Courtois, M., Lanoue, A., Dugé de Bernonville, T., Besseau, S., Glévarec, G., Papon, N., Atehortúa, L., Giglioli-Guivarc'h, N., Clastre, M., St-Pierre, B., Oudin, A., & Courdavault, V. (2017). Virus-

induced gene silencing in *Rauwolfia* species. *Protoplasma*, *254*(4), 1813–1818. https://doi.org/10.1007/s00709-017-1079-y

Courdavault, V., O'Connor, S. E., Jensen, M. K., & Papon, N. (2021). Metabolic engineering for plant natural products biosynthesis: new procedures, concrete achievements and remaining limits. *Natural Product Reports*. https://doi.org/10.1039/D0NP00092B

Courdavault, V., O'Connor, S. E., Oudin, A., Besseau, S., Papon, N., O'Connor, S. E., Oudin, A., Besseau, S., & Papon, N. (2020). Towards the Microbial Production of Plant-Derived Anticancer Drugs. *Trends in Cancer*, *6*(6), 444–448. https://doi.org/10.1016/j.trecan.2020.02.004

Courdavault, V., Papon, N., Clastre, M., Giglioli-Guivarc'h, N., St-Pierre, B., & Burlat, V. (2014). A look inside an alkaloid multisite plant: the *Catharanthus* logistics. *Current Opinion in Plant Biology*, *19*, 43–50. https://doi.org/https://doi.org/10.1016/j.pbi.2014.03.010

D

Dang, T. T. T., & Facchini, P. J. (2012). Characterization of three O-methyltransferases involved in noscapine biosynthesis in opium poppy. *Plant Physiology*, *159*(2), 618–631. https://doi.org/10.1104/pp.112.194886

DeBono, A., Capuano, B., & Scammells, P. J. (2015). Progress toward the development of noscapine and derivatives as anticancer agents. *Journal of Medicinal Chemistry*, *58*(15), 5699–5727. https://doi.org/10.1021/jm501180v

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), 602–607. https://doi.org/10.1038/nature09886

Dey, P., Kundu, A., Kumar, A., Gupta, M., Lee, B. M., Bhakta, T., Dash, S., & Kim, H. S. (2020). Analysis of alkaloids (indole alkaloids, isoquinoline alkaloids, tropane alkaloids). In *Recent Advances in Natural Products Analysis* (pp. 505–567). Elsevier. https://doi.org/10.1016/B978-0-12-816455-6.00015-9

Dugé De Bernonville, T., Carqueijeiro, I., Lanoue, A., Lafontaine, F., Sánchez Bel, P., Liesecke, F., Musset, K., Oudin, A., Glévarec, G., Pichon, O., Besseau, S., Clastre, M., St-Pierre, B., Flors, V., Maury, S., Huguet, E., O'Connor, S. E., & Courdavault, V. (2017). Folivory elicits a strong defense reaction in *Catharanthus roseus*: Metabolomic and transcriptomic analyses reveal distinct local and systemic responses. *Scientific Reports*, *7*:40453. https://doi.org/10.1038/srep40453

Dugé de Bernonville, T., Clastre, M., Besseau, S., Oudin, A., Burlat, V., Glévarec, G., Lanoue, A., Papon, N., Giglioli-Guivarc'h, N., St-Pierre, B., & Courdavault, V. (2015). Phytochemical genomics of the Madagascar periwinkle: unravelling the last twists of the alkaloid engine. *Phytochemistry*, *113*, 9–23. https://doi.org/https://doi.org/10.1016/j.phytochem.2014.07.023

Dugé de Bernonville, T., Papon, N., Clastre, M., O'Connor, S. E., & Courdavault, V. (2020). Identifying missing biosynthesis enzymes of plant natural products. *Trends in Pharmacological Sciences*, *41*(3), 142–146. https://doi.org/10.1016/j.tips.2019.12.006

Dusséaux, S., Wajn, W. T., Liu, Y., Ignea, C., & Kampranis, S. C. (2020). Transforming yeast peroxisomes into microfactories for the efficient production of high-value isoprenoids. *Proceedings of the National Academy of Sciences*, *117*(50), 31789–31799. https://doi.org/10.1073/pnas.2013968117

E

Erhirhie, E. O., Ikegbune, C., Okeke, A. I., Onwuzuligbo, C. C., Madubuogwu, N. U., Chukwudulue, U. M., & Okonkwo, O. B. (2021). Antimalarial herbal drugs: a review of their interactions with conventional antimalarial drugs. *Clinical Phytoscience*, *7*(1), 4. https://doi.org/10.1186/s40816-020-00242-4

F

Favarin, D. C., Oliveira, J. D. R., Oliveira, C. J. F., & Rogerio, A. D. P. (2013). Potential effects of medicinal plants and secondary metabolites on acute lung injury. *BioMed Research International*, *2013*, 576479. https://doi.org/10.1155/2013/576479

Fischer, R., Vaquero-Martin, C., Sack, M., Drossard, J., Emans, N., & Commandeur, U. (1999). Towards molecular farming in the future: transient protein expression in plants. *Biotechnology and Applied Biochemistry*, *30*(2), 113–116. https://doi.org/https://doi.org/10.1111/j.1470-8744.1999.tb00900.x

Frei, M. (2013). Lignin: characterization of a multifaceted crop component. *The Scientific World Journal*, *2013*, 1–25. https://doi.org/10.1155/2013/436517

G

Gelvin, S. B. (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiology and Molecular Biology Reviews*, *67*(1), 16–37. https://doi.org/10.1128/MMBR.67.1.16-37.2003

Gigant, B., Wang, C., Ravelli, R. B. G., Roussi, F., Steinmetz, M. O., Curmi, P. A., Sobel, A., & Knossow, M. (2005). Structural basis for the regulation of tubulin by vinblastine. *Nature*, *435*(7041), 519–522. https://doi.org/10.1038/nature03566

Góngora-Castillo, E., Childs, K. L., Fedewa, G., Hamilton, J. P., Liscombe, D. K., Magallanes-Lundback, M., Mandadi, K. K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., DellaPenna, D., McKnight, T. D., O'Connor, S., & Buell, C. R. (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE*, *7*(12), e52506. https://doi.org/10.1371/journal.pone.0052506

Grigalius, I., & Petrikaite, V. (2017). Relationship between antioxidant and anticancer activity of trihydroxyflavones. *Molecules*, *22*(12), 2169. https://doi.org/10.3390/molecules22122169

Grynkiewicz, G., & Gadzikowska, M. (2008). Tropane alkaloids as medicinally useful natural products and their synthetic derivatives as new drugs. *Pharmacological Reports*, *60*(4), 439–463.

Guirimand, G., Courdavault, V., Lanoue, A., Mahroug, S., Guihur, A., Blanc, N., Giglioli-Guivarc'h, N., St-Pierre, B., & Burlat, V. (2010). Strictosidine activation in Apocynaceae: Towards a "nuclear time bomb"? *BMC Plant Biology*, *10*:182, 1–20. https://doi.org/10.1186/1471-2229-10-182

Guirimand, G., Guihur, A., Poutrain, P., Héricourt, F., Mahroug, S., St-Pierre, B., Burlat, V., & Courdavault, V. (2011). Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *Journal of Plant Physiology*, *168*(6), 549–557. https://doi.org/10.1016/j.jplph.2010.08.018

Guirimand, G., Kulagina, N., Papon, N., Hasunuma, T., & Courdavault, V. (2021). Innovative tools and strategies for optimizing yeast cell factories. *Trends in Biotechnology*, *39*(5), 488–504. https://doi.org/10.1016/j.tibtech.2020.08.010

Gunning, P. W., & Hardeman, E. C. (2018). Isoforms: fundamental differences. *ELife*, *7:* e34477, . https://doi.org/10.7554/eLife.34477

Guo, D., Chen, F., Inoue, K., Blount, J. W., & Dixon, R. A. (2001). Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic Alfalfa: Impacts on lignin structure and implications for the biosynthesis of G and S lignin. *The Plant Cell*, *13*(1), 73–88. https://doi.org/10.1105/tpc.13.1.73

Guo, Q., Liu, Q., A. Smith, N., Liang, G., & Wang, M.-B. (2016). RNA silencing in plants: mechanisms, Technologies and applications in horticultural crops. *Current Genomics*, *17*(6), 476–489. https://doi.org/10.2174/1389202917666160520103117

H

Hagel, J. M., & Facchini, P. J. (2013). Benzylisoquinoline alkaloid metabolism: A century of discovery and a brave new world. *Plant and Cell Physiology*, *54*(5), 647–672. https://doi.org/10.1093/pcp/pct020

Hahlbrock, K., & Scheel, D. (1989). Biology of phenylpropanoid metabolism. *Annual Review of Plant Physiology and Plant Molecular Biology*, *40*, 347–369.

Hallard, D., Van Der Heijden, R., Verpoorte, R., Lopes Cardoso, M. I., Pasquali, G., Memelink, J., & Hoge, J. H. C. (1997). Suspension cultured transgenic cells of *Nicotiana tabacum* expressing tryptophan decarboxylase and strictosidine synthase cDNAs from *Catharanthus roseus* produce strictosidine upon secologanin feeding. *Plant Cell Reports*, *17*(1), 50–54. https://doi.org/10.1007/s002990050350

Hammond, S. M. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, *293*(5532), 1146–1150. https://doi.org/10.1126/science.1064023

Hartmann, T. (2007). From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*, *68*(22–24), 2831–2846. https://doi.org/10.1016/j.phytochem.2007.09.017

Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K., & Mueller-Roeber, B. (2018). L-SCRaMbLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nature Communications*, *9*(1), 1931. https://doi.org/10.1038/s41467-017-02208-6

Holland, H. L., & Weber, H. K. (2000). Enzymatic hydroxylation reactions. *Current Opinion in Biotechnology*, *11*(6), 547–553. https://doi.org/10.1016/S0958-1669(00)00142-7

Hou, Q., He, W. J., Wu, Y. S., Hao, H. J., Xie, X. Y., & Fu, X. B. (2020). Berberine: A traditional natural product with novel biological activities. *Alternative Therapies in Health and Medicine*, *26*(S2), 20–27.

Hussein, R. A., & El-Anssary, A. A. (2019). Plants secondary metabolites: The key drivers of the pharmacological actions of medicinal plants. In *Herbal Medicine*. IntechOpen. https://doi.org/10.5772/intechopen.76139

I

Ibdah, M., Zhang, X. H., Schmidt, J., & Vogt, T. (2003). A novel Mg2+-dependent O-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. *Journal of Biological Chemistry*, *278*(45), 43961–43972. https://doi.org/10.1074/jbc.M304932200

Ibrahim, R. K., Bruneau, A., & Bantignies, B. (1998). Plant O-methyltransferases: Molecular analysis, common signature and classification. *Plant Molecular Biology*, *36*(1), 1–10. https://doi.org/10.1023/A:1005939803300

Ibrahim, R. K., De Luca, V., Khouri, H., Latchinian, L., Brisson, L., & Charest, P. M. (1987). Enzymology and compartmentation of polymethylated flavonol glucosides in *Chrysosplenium Americanum*. *Phytochemistry*, *26*(5), 1237–1245. https://doi.org/10.1016/S0031-9422(00)81789-6

Ibrahim, R. K., & Muzac, I. (2000). The methyltransferase gene superfamily: a tree with multiple branches. *Recent Advances in Phytochemistry*.

Ishitani, M., Majumder, A. L., Bornhouser, A., Michalowski, C. B., Jensen, R. G., & Bohnert, H. J. (1996). Coordinate transcriptional induction of myo-inositol metabolism during environmental stress. *The Plant Journal*, *9*(4), 537–548. https://doi.org/10.1046/j.1365-313X.1996.09040537.x

Itoh, N., Toda, H., Matsuda, M., Negishi, T., Taniguchi, T., & Ohsawa, N. (2009). Involvement of S-adenosylmethionine-dependent halide/thiol methyltransferase (HTMT) in methyl halide emissions from agricultural plants: Isolation and characterization of an HTMT-coding gene from *Raphanus sativus* (daikon radish). *BMC Plant Biology*, *9*, 1–10. https://doi.org/10.1186/1471-2229-9-116

Iwakiri, T., Mase, S., Murakami, T., Matsumoto, M., Hamada, H., Nakayama, T., & Ozaki, S. I. (2013). Glucosylation of hydroxyflavones by glucosyltransferases from *Phytolacca americana*. *Journal of Molecular Catalysis B: Enzymatic*, *90*, 61–65. https://doi.org/10.1016/j.molcatb.2013.01.016

J

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166. https://doi.org/10.1126/science.860134

Jeon, Y. M., Kim, B. G., & Ahn, J. H. (2009). Biological synthesis of 7-O-methyl apigenin from naringenin using *Escherichia coli* expressing two genes. *Journal of Microbiology and Biotechnology*, *19*(5), 491–494. https://doi.org/10.4014/jmb.0807.402

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable Dual-RNA-Guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jogaiah, S., & Abdelrahman, M. (Eds.). (2019). *Bioactive Molecules in Plant Defense*. Springer International Publishing. https://doi.org/10.1007/978-3-030-27165-7

Jones, S. F., & Burris, H. A. (1996). Vinorelbine: A new antineoplastic drug for the treatment of non-small-cell lung cancer. *Annals of Pharmacotherapy*, *30*(5), 501–506. https://doi.org/10.1177/106002809603000513

Joshi, C. P., & Chiang, V. L. (1998). Conserved sequence motifs in plant S-adenosyl-L-methionine-dependent methyltransferases. *Plant Molecular Biology*, *37*(4), 663–674. https://doi.org/10.1023/A:1006035210889

K

Kapila, J., De Rycke, R., Van Montagu, M., & Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant Science*, *122*(1), 101–108. https://doi.org/10.1016/S0168-9452(96)04541-4

Karlsson, M. O., Dahlström, B., & Neil, A. (1988). Characterization of high-affinity binding sites for the antitussive [3H] noscapine in guinea pig brain tissue. *European Journal of Pharmacology*, *145*(2), 195–203. https://doi.org/10.1016/0014-2999(88)90230-0

Khalid, M., Saeed-ur-Rahman, Bilal, M., & Huang, D. (2019). Role of flavonoids in plant interactions with the environment and against human pathogens — A review. *Journal of Integrative Agriculture*, *18*(1), 211–230. https://doi.org/10.1016/S2095-3119(19)62555-4

Khalil, M., Brandt, W., Beuerle, T., Reckwell, D., Groeneveld, J., Hansch, R., Gaid, M., Liu, B., & Beerhues, L. (2015). O-Methyltransferases involved in biphenyl and dibenzofuran biosynthesis. *Plant J*, *83*(2), 263–276. https://doi.org/10.1111/tpj.12885

Kilgore, M. B., Augustin, M. M., Starks, C. M., O'Neil-Johnson, M., May, G. D., Crow, J. A., & Kutchan, T. M. (2014). Cloning and characterization of a norbelladine 4′-O-methyltransferase involved in the biosynthesis of the alzheimer's drug galanthamine in *Narcissus* sp. aff. *pseudonarcissus*. *PLoS ONE*, *9*(7), e103223. https://doi.org/10.1371/journal.pone.0103223

Kim, B. G., Sung, S. H., Chong, Y., Lim, Y., & Ahn, J. H. (2010). Plant flavonoid O-methyltransferases: substrate specificity and application. *Journal of Plant Biology*, *53*(5), 321–329. https://doi.org/10.1007/s12374-010-9126-7

Kim, G. B., Kim, W. J., Kim, H. U., & Lee, S. Y. (2020). Machine learning applications in systems metabolic engineering. *Current Opinion in Biotechnology*, *64*, 1–9. https://doi.org/10.1016/j.copbio.2019.08.010

Koudounas, K., Guirimand, G., Hoyos, L. F. R., Carqueijeiro, I., Cruz, P. L., Stander, E., Kulagina, N., Perrin, J., Oudin, A., Besseau, S., Lanoue, A., Atehortùa, L., St-Pierre, B., Giglioli-Guivarc'h, N., Papon, N., O'Connor, S. E. & Vincent Courdavault (2021). Tonoplast and peroxisome targeting of γ-tocopherol *N*-methyltransferase homologs involved in the synthesis of monoterpene indole alkaloids. *Manuscript submitted for publication*

Kulagina, N., Besseau, S., Papon, N., & Courdavault, V. (2021). Peroxisomes: a new hub for metabolic engineering in yeast. *Frontiers in Bioengineering and Biotechnology*, *9*, 207. https://doi.org/10.3389/fbioe.2021.659431

Kumar, S., Singh, A., Bajpai, V., Srivastava, M., Singh, B. P., & Kumar, B. (2016). Structural characterization of monoterpene indole alkaloids in ethanolic extracts of *Rauwolfia* species by liquid chromatography with quadrupole time-of-flight mass spectrometry. *Journal of Pharmaceutical Analysis*, *6*(6), 363–373. https://doi.org/10.1016/j.jpha.2016.04.008

Kurek, J. (2019). Introductory Chapter: Alkaloids - their importance in nature and for human Life. In *Alkaloids - their importance in nature and human life*. IntechOpen. https://doi.org/10.5772/intechopen.85400

Kusano, H., Li, H., Minami, H., Kato, Y., Tabata, H., & Yazaki, K. (2019). Evolutionary developments in plant specialized metabolism, exemplified by two transferase families. *Frontiers in Plant Science*, *10*, 794. https://doi.org/10.3389/fpls.2019.00794

L

Lam, K. C., Ibrahim, R. K., Behdad, B., & Dayanandan, S. (2007). Structure, function, and evolution of plant O-methyltransferases. *Genome*, *50*(11), 1001–1013. https://doi.org/10.1139/G07-077

Lapeña, D., Kosa, G., Hansen, L.D., Mydland, L.T., Passoth, V., Horn, S.J. & Eijsink, V.G.H. (2020). Production and characterization of yeasts grown on media composed of spruce-derived sugars and protein hydrolysates from chicken by-products. *Microbial Cell Factories*, *19*(19). https://doi.org/10.1186/s12934-020-1287-6.

Larsen, B., Fuller, V. L., Pollier, J., Van Moerkercke, A., Schweizer, F., Payne, R., Colinas, M., O'Connor, S. E., Goossens, A., & Halkier, B. A. (2017). Identification of iridoid glucoside transporters in *Catharanthus roseus*. *Plant and Cell Physiology*, *58*(9), 1507–1518. https://doi.org/10.1093/pcp/pcx097

Levac, D., Murata, J., Kim, W. S., & De Luca, V. (2008). Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-O-methyltransferase. *Plant Journal*, *53*(2), 225–236. https://doi.org/10.1111/j.1365-313X.2007.03337.x

Lewis, N. G., & Yamamoto, E. (1990). Lignin: Occurrence, biogenesis and biodegradation. *Annual Review of Plant Physiology and Plant Molecular Biology*, *41*(1), 455–496. https://doi.org/10.1146/annurev.pp.41.060190.002323

Li, L., Christian J. Stoeckert, Jr., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189. https://doi.org/10.1101/gr.1224503

Li, Y., Kong, D., Fu, Y., Sussman, M. R., & Wu, H. (2020). The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiology and Biochemistry*, *148*, 80–89. https://doi.org/https://doi.org/10.1016/j.plaphy.2020.01.006

Li, Y., & Smolke, C. D. (2016). Engineering biosynthesis of the anticancer alkaloid noscapine in yeast. *Nature Communications*, *7*(1), 12137. https://doi.org/10.1038/ncomms12137

Li, Y. Y., Chen, S. W., Yang, L. M., Wang, R. R., Pang, W., & Zheng, Y. T. (2009). The anti-HIV actions of 7-and 10-substituted camptothecins. *Molecules*, *15*(1), 138–148. https://doi.org/10.3390/molecules15010138

Liscombe, D. K., Usera, A. R., & O'Connor, S. E. (2010). Homolog of tocopherol *C*-methyltransferases catalyzes *N*-methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences*, *107*(44), 18793–18798. https://doi.org/10.1073/pnas.1009003107

Liu, C. J., Deavours, B. E., Richard, S. B., Ferrer, J. L., Blount, J. W., Huhman, D., Dixon, R. A., & Noel, J. P. (2006). Structural basis for dual functionality of isoflavonoid O-methyltransferases in the evolution of plant defense responses. *Plant Cell*, *18*(12), 3656–3669. https://doi.org/10.1105/tpc.106.041376

Lobay, D. (2015). *Rauwolfia* in the treatment of hypertension. *Integrative Medicine (Encinitas, Calif.)*, *14*(3), 40–46.

Luo, X., Reiter, M. A., D'Espaux, L., Wong, J., Denby, C. M., Lechner, A., Zhang, Y., Grzybowski, A. T., Harth, S., Lin, W., Lee, H., Yu, C., Shin, J., Deng, K., Benites, V. T., Wang, G., Baidoo, E. E. K., Chen, Y., Dev, I., Keasling, J. D. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature*, *567*(7746), 123–126. https://doi.org/10.1038/s41586-019-0978-9

M

Marles, R. J., & Farnsworth, N. R. (1995). Antidiabetic plants and their active constituents. *Phytomedicine*, *2*(2), 137–189. https://doi.org/10.1016/S0944-7113(11)80059-0

Matsuura, H. N., & Fett-Neto, A. G. (2015). Plant alkaloids: main features, toxicity, and mechanisms of action. In *Plant Toxins* (pp. 1–15). Springer Netherlands. https://doi.org/10.1007/978-94-007-6728-7_2-1

Maury, S., Geoffroy, P., & Legrand, M. (1999). Tobacco O-methyltransferases involved in phenylpropanoid metabolism. The different caffeoyl-coenzyme A/5-hydroxyferuloyl-coenzyme A 3/5-O-methyltransferase and caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase classes have distinct substrate specificities and expression patterns. *Plant Physiology*, *121*(1), 215–224. https://doi.org/10.1104/pp.121.1.215

Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K.-M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J., & Werck-Reichhart, D. (2014). The seco-iridoid pathway from *Catharanthus roseus*. *Nature Communications*, *5*(1), 3606. https://doi.org/10.1038/ncomms4606

Miyazaki, T., Yokoshima, S., Simizu, S., Osada, H., Tokuyama, H., & Fukuyama, T. (2007). Synthesis of (+)-vinblastine and its analogues. *Organic Letters*, *9*(23), 4737–4740. https://doi.org/10.1021/ol702040y

Montanher, A. B., Zucolotto, S. M., Schenkel, E. P., & Fröde, T. S. (2007). Evidence of anti-inflammatory effects of *Passiflora edulis* in an inflammation model. *Journal of Ethnopharmacology*, *109*(2), 281–288. https://doi.org/10.1016/j.jep.2006.07.031

Morishige, T., Dubouzet, E., Choi, K. B., Yazaki, K., & Sato, F. (2002). Molecular cloning of columbamine O-methyltransferase from cultured *Coptis japonica* cells. *European Journal of Biochemistry*, *269*(22), 5659–5667. https://doi.org/10.1046/j.1432-1033.2002.03275.x

Morishige, T., Tamakoshi, M., Takemura, T., & Sato, F. (2010). Molecular characterization of O-methyltransferases involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, *86*(7), 757–768. https://doi.org/10.2183/pjab.86.757

Morishige, T., Tsujita, T., Yamada, Y., & Sato, F. (2000). Molecular characterization of the S-Adenosyl-l-methionine:3′-Hydroxy-N-methylcoclaurine 4′-O-methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *Journal of Biological Chemistry*, *275*(30), 23398–23405. https://doi.org/10.1074/jbc.M002439200

Morris, J. S., & Facchini, P. J. (2019). Molecular origins of functional diversity in benzylisoquinoline alkaloid methyltransferases. *Frontiers in Plant Science, 10*, 1–22. https://doi.org/10.3389/fpls.2019.01058

Murata, J., Roepke, J., Gordon, H., & De Luca, V. (2008). The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *The Plant Cell*, *20*(3), 524–542. https://doi.org/10.1105/tpc.107.056630

N

Nakayama, T. (2000). Aureusidin synthase: A polyphenol oxidase homolog responsible for flower coloration. *Science*, *290*(5494), 1163–1166. https://doi.org/10.1126/science.290.5494.1163

Nett, R. S., Lau, W., & Sattely, E. S. (2020). Discovery and engineering of colchicine alkaloid biosynthesis. *Nature*, *584*(7819), 148–153. https://doi.org/10.1038/s41586-020-2546-8

Noble, R. L. (1990). The discovery of the *vinca* alkaloids—chemotherapeutic agents against cancer. *Biochemistry and Cell Biology*, *68*(12), 1344–1351. https://doi.org/10.1139/o90-197

Nomura, T., & Kutchan, T. M. (2010). Three new O-methyltransferases are sufficient for all O-methylation reactions of ipecac alkaloid biosynthesis in root culture of *Psychotria ipecacuanha*. *Journal of Biological Chemistry*, *285*(10), 7722–7738. https://doi.org/10.1074/jbc.M109.086157

O

O'Connor, S. E. (2012). Strategies for engineering plant natural products. In *Methods in Enzymology* (pp. 189–206). https://doi.org/10.1016/B978-0-12-394290-6.00009-4

Ounaroon, A., Decker, G., Schmidt, J. J., Lottspeich, F., & Kutchan, T. M. (2003). (R,S)-Reticuline 7-O-methyltransferase and (R,S)-norcoclaurine 6-*O*-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. *Plant J*, *36*(6), 808–819. https://doi.org/10.1046/j.1365-313X.2003.01928.x

P

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Jiang, H., Fleck, M., Bajad, S., Dang, G., Newman, J. D. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, *496*(7446), 528–532. https://doi.org/10.1038/nature12051

Park, M. R., Chen, X., Lang, D. E., Ng, K. K. S., & Facchini, P. J. (2018). Heterodimeric O-methyltransferases involved in the biosynthesis of noscapine in opium poppy. *The Plant Journal*, *95*(2), 252–267. https://doi.org/10.1111/tpj.13947

Pathak, S., Lakhwani, D., Gupta, P., Mishra, B. K., Shukla, S., Asif, M. H., & Trivedi, P. K. (2013). Comparative transcriptome analysis using high papaverine mutant of *Papaver somniferum* reveals pathway and uncharacterized steps of papaverine biosynthesis. *PLoS ONE*, *8*(5), e65622. https://doi.org/10.1371/journal.pone.0065622

Patil, N. D., Tanguy, N. R., & Yan, N. (2016). Lignin interunit linkages and model compounds. In *Lignin in Polymer Composites* (pp. 27–47). Elsevier. https://doi.org/10.1016/B978-0-323-35565-0.00003-5

Payne, R. M. E., Xu, D., Foureau, E., Teto Carqueijeiro, M. I. S., Oudin, A., Bernonville, T. D. de, Novak, V., Burow, M., Olsen, C.-E., Jones, D. M., Tatsis, E. C., Pendle, A., Ann Halkier, B., Geu-Flores, F., Courdavault, V., Nour-Eldin, H. H., & O'Connor, S. E. (2017). An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nature Plants*, *3*(2), 16208. https://doi.org/10.1038/nplants.2016.208

Peng, M., Shahzad, R., Gul, A., Subthain, H., Shen, S., Lei, L., Zheng, Z., Zhou, J., Lu, D., Wang, S., Nishawy, E., Liu, X., Tohge, T., Fernie, A. R., & Luo, J. (2017). Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nature Communications*, *8*(1), 1975. https://doi.org/10.1038/s41467-017-02168-x

Pflieger, S., Richard, M. M. S., Blanchet, S., Meziadi, C., & Geffroy, V. (2013). VIGS technology: an attractive tool for functional genomics studies in legumes. *Functional Plant Biology*, *40*(12), 1234. https://doi.org/10.1071/FP13089

Pickens, L. B., Tang, Y., & Chooi, Y. H. (2011). Metabolic engineering for the production of natural products. *Annual Review of Chemical and Biomolecular Engineering*, *2*, 211-236. https://doi.org/10.1146/annurev-chembioeng-061010-114209

Pienkny, S., Brandt, W., Schmidt, J., Kramell, R., & Ziegler, J. (2009). Functional characterization of a novel benzylisoquinoline O-methyltransferase suggests its involvement in papaverine biosynthesis in opium poppy (*Papaver somniferum L*). *The Plant Journal, 60*, 56-67. https://doi.org/10.1111/j.1365-313X.2009.03937.x

Pietrocola, F., Galluzzi, L., Bravo-San Pedro, J. M., Madeo, F., & Kroemer, G. (2015). Acetyl coenzyme A: A central metabolite and second messenger. *Cell Metabolism*, *21*(6), 805–821. https://doi.org/10.1016/j.cmet.2015.05.014

Pott, D. M., Osorio, S., & Vallarino, J. G. (2019). From central to specialized metabolism: an overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. *Frontiers in Plant Science*, *10*:835. https://doi.org/10.3389/fpls.2019.00835

Proksa, B., Uhrín, D., Grossmann, E., Votický, Z., & Fuska, J. (1988). Relative configuration and cytotoxic activity of vincarubine: a novel bisindole alkaloid from *Vinca minor*. *Planta Medica*, *54*(03), 214–218. https://doi.org/10.1055/s-2006-962407

Pyne, M.E., Kevvai, K., Grewal, P.S., Narcross, L., Choi, B., Bourgeois, L., Dueber, J.E. & Martin, V.J.J. (2020). A yeast platform for high-level synthesis of tetrahydroisoquinoline alkaloids. *Nature Communications*, *11*, 1–10. https://doi.org/10.1038/s41467-020-17172-x.

Q

Qu, Y., Easson, M. L. A. E., Froese, J., Simionescu, R., Hudlicky, T., & De Luca, V. (2015). Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proceedings of the National Academy of Sciences*, *112*(19), 6224–6229. https://doi.org/10.1073/pnas.1501821112

R

Ratcliff, F. G., MacFarlane, S. A., & Baulcombe, D. C. (1999). Gene silencing without DNA: RNA-Mediated cross-protection between viruses. *The Plant Cell*, *11*(7), 1207–1215. https://doi.org/10.1105/tpc.11.7.1207

Riley, L. A., & Guss, A. M. (2021). Approaches to genetic tool development for rapid domestication of non-model microorganisms. *Biotechnology for Biofuels*, *14*(1), 30. https://doi.org/10.1186/s13068-020-01872-z

Roberts M.F. (1998) Enzymology of alkaloid biosynthesis. In: Roberts M.F., Wink M. (eds) *Alkaloids*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4757-2905-4_5

Robin, A. Y., Giustini, C., Graindorge, M., Matringe, M., & Dumas, R. (2016). Crystal structure of norcoclaurine-6-O-methyltransferase, a key rate-limiting step in the synthesis of benzylisoquinoline alkaloids. *The Plant Journal*, *87*(6), 641–653. https://doi.org/10.1111/tpj.13225

Rodriguez-Trelles, F., Tarrio, R., & Ayala, F. J. (2003). Convergent neofunctionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. *Proceedings of the National Academy of Sciences*, *100*(23), 13413–13417. https://doi.org/10.1073/pnas.1835646100

S

Sainsbury, F., & Lomonossoff, G. P. (2008). Extremely high-level and rapid transient protein production in plants without the use of viral replication. *Plant Physiology*, *148*(3), 1212–1218. https://doi.org/10.1104/pp.108.126284

Sainsbury, F., Thuenemann, E. C., & Lomonossoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal*, *7*(7), 682–693. https://doi.org/10.1111/j.1467-7652.2009.00434.x

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., & Fernie, A. R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: Structural and genetic diversity. *Plant Physiology and Biochemistry*, *72*, 21–34. https://doi.org/10.1016/j.plaphy.2013.02.001

Sajja, R. K., Rahman, S., & Cucullo, L. (2016). Drugs of abuse and blood-brain barrier endothelial dysfunction: A focus on the role of oxidative stress. *Journal of Cerebral Blood Flow & Metabolism*, *36*(3), 539–554. https://doi.org/10.1177/0271678X15616978

Salim, V., Jones, A. D., DellaPenna, D., Jones, D., & DellaPenna, D. (2018). *Camptotheca acuminata* 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism. *Plant Journal*, *95*(1), 112–125. https://doi.org/10.1111/tpj.13936

Samanta, A., Das, G., & Das, S. (2011). Roles of flavonoids in plants. *Int J Pharm Sci Tech*, 6(1), 12-35. ISSN: 0975-0525

Sato, F., Tsujita, T., Katagiri, Y., Yoshida, S., & Yamada, Y. (1994). Purification and characterization of S-adenosyl-methionine : norcoclaurine 6-O-methyltransferase from cultured *Coptis japonica* cells. *European Journal of Biochemistry*, *225*(1), 125–131. https://doi.org/10.1111/j.1432-1033.1994.00125.x

Schröder, G., Wehinger, E., Lukačin, R., Wellmann, F., Seefelder, W., Schwab, W., & Schröder, J. (2004). Flavonoid methylation: a novel 4′-O-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. *Phytochemistry*, *65*(8), 1085–1094. https://doi.org/10.1016/j.phytochem.2004.02.010

Sengupta, S., Patra, B., Ray, S., & Majumder, A. L. (2008). Inositol methyl transferase from a halophytic wild rice, *Porteresia coarctata Roxb.* (Tateoka): regulation of pinitol synthesis

under abiotic stress. *Plant, Cell & Environment*, *31*(10), 1442–1459. https://doi.org/10.1111/j.1365-3040.2008.01850.x

Simon, C., Lion, C., Biot, C., Gierlinger, N., & Hawkins, S. (2018). Lignification and advances in lignin imaging in plant cell walls. In *Annual Plant Reviews online* (pp. 909–940). Wiley. https://doi.org/10.1002/9781119312994.apr0656

Singh, S., McCoy, J. G., Zhang, C., Bingman, C. A., Phillips, G. N., & Thorson, J. S. (2008). Structure and mechanism of the rebeccamycin sugar 4′-O-methyltransferase RebM. *Journal of Biological Chemistry*, *283*(33), 22628–22636. https://doi.org/10.1074/jbc.M800503200

Srinivasan, P., & Smolke, C. D. (2019). Engineering a microbial biosynthesis platform for *de novo* production of tropane alkaloids. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-11588-w

Srivastava, V., Negi, A. S., Kumar, J. K., Gupta, M. M., & Khanuja, S. P. S. (2005). Plant-based anticancer molecules: A chemical and biological profile of some important leads. *Bioorganic & Medicinal Chemistry*, *13*(21), 5892–5908. https://doi.org/10.1016/j.bmc.2005.05.066

Stermitz, F. R. (2000). Poppy. The Genus *Papaver*. Harwood Academic Publishers, Amsterdam, The Netherlands. *Journal of Natural Products*, *63*(1), 169–170. https://doi.org/10.1021/np990766g

St-Pierre, B., & De Luca, V. (1995). A cytochrome P-450 monooxygenase catalyzes the first step in the conversion of tabersonine to vindoline in *Catharanthus roseus*. *Plant Physiology*, *109*(1), 131–139. https://doi.org/10.1104/pp.109.1.131

Sung, S. H., Kim, B. G., & Ahn, J. H. (2011). Optimization of rhamnetin production in *Escherichia coli*. *Journal of Microbiology and Biotechnology*, *21*(8), 854–857. https://doi.org/10.4014/jmb.1104.04048

Sung, Y.-C., Lin, C.-P., & Chen, J.-C. (2014). Optimization of virus-induced gene silencing in *Catharanthus roseus*. *Plant Pathology*, *63*(5), 1159–1167. https://doi.org/10.1111/ppa.12186

Szczebara, F. M., Chandelier, C., Villeret, C., Masurel, A., Bourot, S., Duport, C., Blanchard, S., Groisillier, A., Testet, E., Costaglioli, P., Cauet, G., Degryse, E., Balbuena, D., Winter, J., Achstetter, T., Spagnoli, R., Pompon, D., & Dumas, B. (2003). Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature Biotechnology*, *21*(2), 143–149. https://doi.org/10.1038/nbt775

T

Takshak, S., & Agrawal, S. B. (2019). Defense potential of secondary metabolites in medicinal plants under UV-B stress. *Journal of Photochemistry and Photobiology B: Biology*, *193*, 51–88. https://doi.org/https://doi.org/10.1016/j.jphotobiol.2019.02.002

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, *408*(6814), 796–815. https://doi.org/10.1038/35048692

Tohge, T., & Fernie, A. R. (2017). An overview of compounds derived from the shikimate and phenylpropanoid pathways and their medicinal importance. *Mini-Reviews in Medicinal Chemistry*, *17*(12). https://doi.org/10.2174/1389557516666160624123425

Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., & Fernie, A. R. (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in *Brassicaceae*. *Nature Communications*, *7*(1), 12399. https://doi.org/10.1038/ncomms12399

Tzfira, T., & Citovsky, V. (2006). Agrobacterium-mediated genetic transformation of plants: biology and biotechnology. *Current Opinion in Biotechnology*, *17*(2), 147–154. https://doi.org/10.1016/j.copbio.2006.01.009

U

Ullrich, R., & Hofrichter, M. (2007). Enzymatic hydroxylation of aromatic compounds. *Cellular and Molecular Life Sciences*, *64*(3), 271–293. https://doi.org/10.1007/s00018-007-6362-1

V

Van der Heijden, R., Jacobs, D., Snoeijer, W., Hallard, D., & Verpoorte, R. (2004). The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Current Medicinal Chemistry*, *11*(5), 607–628. https://doi.org/10.2174/0929867043455846

Vaquero, C., Sack, M., Chandler, J., Drossard, J., Schuster, F., Monecke, M., Schillberg, S., & Fischer, R. (1999). Transient expression of a tumor-specific single-chain fragment and a chimeric antibody in tobacco leaves. *Proceedings of the National Academy of Sciences*, *96*(20), 11128–11133. https://doi.org/10.1073/pnas.96.20.11128

Vázquez-Flota, F., Carrillo-Pech, M., Minero-García, Y., & de Lourdes Miranda-Ham, M. (2004). Alkaloid metabolism in wounded *Catharanthus roseus* seedlings. *Plant Physiology and Biochemistry*, *42*(7–8), 623–628. https://doi.org/10.1016/j.plaphy.2004.06.010

Voinnet, O., Lederer, C., & Baulcombe, D. C. (2000). A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell*, *103*(1), 157–167. https://doi.org/10.1016/S0092-8674(00)00095-7

Voinnet, O., Pinto, Y. M., & Baulcombe, D. C. (1999). Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences*, *96*(24), 14147–14152. https://doi.org/10.1073/pnas.96.24.14147

Voinnet, O., Rivas, S., Mestre, P., & Baulcombe, D. (2003). Retracted: An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *The Plant Journal*, *33*(5), 949–956. https://doi.org/10.1046/j.1365-313X.2003.01676.x

W

Wen, W., Alseekh, S., & Fernie, A. R. (2020). Conservation and diversification of flavonoid metabolism in the plant kingdom. *Current Opinion in Plant Biology*, *55*, 100–108. https://doi.org/10.1016/j.pbi.2020.04.004

Wen, X., & Walle, T. (2006). Methylation protects dietary flavonoids from rapid hepatic metabolism. *Xenobiotica*, *36*(5), 387–397. https://doi.org/10.1080/00498250600630636

Wessjohann, L., Dippe, M., Tengg, M., & Gruber-Khadjawi, M. (2014). Methyltransferases in biocatalysis. In *Cascade Biocatalysis* (eds S. Riva and W. D. Fessner). https://doi.org/10.1002/9783527682492.ch18

Whitham, S. A., Lincoln, L. M., Chowda-Reddy, R. V., Dittman, J. D., O'Rourke, J. A., & Graham, M. A. (2016). Virus-induced gene silencing and transient gene expression in soybean (glycine max) using bean pod mottle virus infectious clones. *Current Protocols in Plant Biology*, *1*(2), 263–283. https://doi.org/10.1002/cppb.20012

Wilson, M. A., Rimando, A. M., & Wolkow, C. A. (2008). Methoxylation enhances stilbene bioactivity in *Caenorhabditis elegans*. *BMC Pharmacology*, *8*, 1–11. https://doi.org/10.1186/1471-2210-8-15

Wink, M. (2008). Evolution of secondary plant metabolism. In *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0001922.pub2

X

Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C. J. S., Chakrabarty, R., Desgagné-Penix, I., Haslam, T. M., Kim, Y.-B., Liu, E., MacNevin, G., Masada-Atsumi, S., Reed, D. W., Stout, J. M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P. S., De Luca, V., Sensen, C. W. (2013). Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *Journal of Biotechnology*, *166*(3), 122–134. https://doi.org/10.1016/j.jbiotec.2013.04.004

Y

Yamamoto, K., Takahashi, K., Caputi, L., Mizuno, H., Rodriguez-Lopez, C. E., Iwasaki, T., Ishizaki, K., Fukaki, H., Ohnishi, M., Yamazaki, M., Masujima, T., O'Connor, S. E., & Mimura, T. (2019). The complexity of intercellular localisation of alkaloids revealed by single-cell metabolomics. *New Phytologist*, *224*(2), 848–859. https://doi.org/10.1111/nph.16138

Yan, Y., Acevedo, M., Mignacca, L., Desjardins, P., Scott, N., Imane, R., Quenneville, J., Robitaille, J., Feghaly, A., Gagnon, E., Ferbeyre, G., & Major, F. (2018). The sequence features that define efficient and specific hAGO2-dependent miRNA silencing guides. *Nucleic Acids Research*, *46*(16), 8181–8196. https://doi.org/10.1093/nar/gky546

Yang, L., Wen, K. S., Ruan, X., Zhao, Y. X., Wei, F., & Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules*, *23*(4), 762. https://doi.org/10.3390/molecules23040762

Yarnell, E. (2002). Botanical medicines for the urinary tract. *World Journal of Urology*, *20*(5), 285–293. https://doi.org/10.1007/s00345-002-0293-0

Yonekura-Sakakibara, K., & Hanada, K. (2011). An evolutionary view of functional diversity in family 1 glycosyltransferases. *The Plant Journal*, *66*(1), 182–193. https://doi.org/https://doi.org/10.1111/j.1365-313X.2011.04493.x

Yun, B. W., Yan, Z., Amir, R., Hong, S., Jin, Y. W., Lee, E. K., & Loake, G. J. (2012). Plant natural products: History, limitations and the potential of cambial meristematic cells. *Biotechnology and Genetic Engineering Reviews*, *28*, 47–60. https://doi.org/10.5661/bger-28-47

Z

Zheng, L., Yang, J., Chen, Y., Ding, L., Wei, J., & Wang, H. (2021). An improved and efficient method of Agrobacterium syringe infiltration for transient transformation and its application in the elucidation of gene function in poplar. *BMC Plant Biology*, *21*(1), 54. https://doi.org/10.1186/s12870-021-02833-w

Zhong, R., Morrison, W. H., Himmelsbach, D. S., Poole, F. L., & Ye, Z. H. (2000). Essential role of caffeoyl coenzyme A O-methyltransferase in lignin biosynthesis in woody poplar plants. *Plant Physiology*, *124*(2), 563–577. https://doi.org/10.1104/pp.124.2.563

Ziegler, J., Diaz-Chávez, M. L., Kramell, R., Ammer, C., & Kutchan, T. M. (2005). Comparative macroarray analysis of morphine containing *Papaver somniferum* and eight morphine free *Papaver* species identifies an O-methyltransferase involved in benzylisoquinoline biosynthesis. *Planta*, *222*(3), 458–471. https://doi.org/10.1007/s00425-005-1550-4

Pamela Lemos Cruz

Elucidation et ingénierie de la biosynthèse d'alcaloïdes d'intérêt pharmaceutique originaires des Apocynacées

Résumé

La pervenche de Madagascar (*Catharanthus roseus*) est une plante de la famille des Apocynacées produisant des alcaloïdes indoliques monoterpéniques (AIM) dont certains présentent un intérêt thérapeutique majeur. *C. roseus* est considérée aujourd'hui comme une plante médicinale modèle en devenant la référence en matière d'étude d'un métabolisme spécialisé, celui des AIM. L'avancée des connaissances sur ce métabolisme a été impulsée par les développements de ressources transcriptomiques alliés à des méthodes de caractérisation fonctionnelle des gènes. L'expertise acquise sur *C. roseus* a également permis la transposition de ces connaissances à d'autres plantes de la famille des Apocynacées. L'objectif de la thèse a consisté à étudier des enzymes de la famille des *O*-méthyltransférases du métabolisme alcaloïdique dans le but de les valoriser dans des approches d'ingénierie métabolique. Ainsi, suite à l'exploitation de données omiques, un gène codant une nouvelle isoforme enzymatique de la tabersonine-16-O-méthyltransférase (16OMT2) a été isolé chez *C. roseus* ainsi qu'un gène orthologue *Vm16OMT* chez l'espèce *Vinca minor*. La validation fonctionnelle du gène *16OMT2* a été réalisée par des études de pertes de fonction grâce à la technique VIGS (Virus-Induced Gene Silencing) et par des études d'expression transitoire utilisant la méthode d'agroinfiltration. Par ailleurs, les deux enzymes ont été caractérisées sur le plan biochimique et leur localisation subcellulaire déterminée. Des recherches en ingénierie métabolique de la levure *Saccharomyces cerevisiae* utilisant les gènes *16OMT2* et *Vm16OMT* ont été menées en vue d'optimiser les premières étapes de bioconversion de la tabersonine en vindoline (l'un des deux précurseurs des alcaloïdes à activité anticancéreuse). Au final, en exploitant les résultats obtenus avec le gène *16OMT2* et d'autres résultats antérieurs obtenus par l'équipe, ces travaux ont conduit à produire en bioréacteur 266 mg/L de vindoline ouvrant la voie à une future production industrielle de ce métabolite.

Mots-clefs: Apocynaceae, *Catharanthus roseus*, alcaloïdes indoliques monoterpéniques, biosynthèse, ingénierie métabolique.

Summary

The Madagascar periwinkle (*Catharanthus roseus*) is a model medicinal plant of the Apocynaceae family producing indole monoterpene alkaloids (MIA) which became the reference in the study of specialized metabolism. The increased knowledge on this metabolism has been driven by the development of transcriptomic resources combined with new methods of functional characterization of genes, which also allowed the transposition of this knowledge to other plants of the Apocynaceae family. The objective of the thesis was to study the *O*-methyltransferase family enzymes involved in the alkaloid metabolism, in order to develop them in metabolic engineering approaches. Thus, following the study of the omics data, a gene encoding a new enzyme isoform of tabersonine-16-*O*-methyltransferase (16OMT2) was isolated in *C. roseus*, as well as an ortholog gene *Vm16OMT* in the *Vinca minor* species. Functional validation of the *16OMT2* gene was carried out by loss-of-function studies using the VIGS (Virus-Induced Gene Silencing) technique and by transient expression studies using the agroinfiltration method. In addition, the two enzymes were characterized biochemically and their subcellular localization described. The metabolic engineering of the yeast *Saccharomyces cerevisiae* was carried out using the *16OMT2* and *Vm16OMT* genes in order to optimize the first steps of tabersonine to vindoline bioconversion (one of the two precursors of alkaloids with anticancer activity). Finally, by exploiting the results obtained with the *16OMT2* gene and other previous results obtained by the team, this work led to the production of 266 mg/L of vindoline in a bioreactor, paving the way for future industrial production of this metabolite.

Mots-clefs: Apocynaceae, *Catharanthus roseus*, monoterpene indole alkaloids, biosynthesis, metabolic engineering.